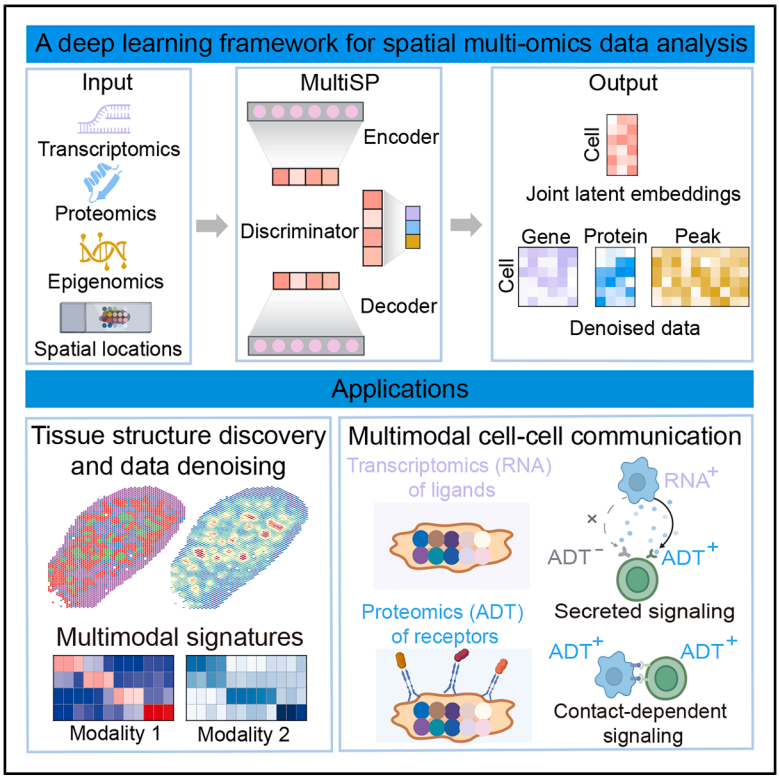
Cell Genomics

Technology

MultiSP deciphers tissue structure and multicellular communication from spatial multi-omics data

Graphical abstract



Highlights

- MultiSP deciphers complex tissue structures from spatial multi-omics data
- Flexible integration of any number of omics modalities
- MultiSP restores noisy and sparse spatial omics data
- Inference of spatially multimodal cell-cell communication

Authors

Chenfeng Mo, Xiufen Zou, Suoqin Jin

Correspondence

sqjin@whu.edu.cn

In brief

Mo et al. present MultiSP, an efficient deep learning framework for integrative analysis of spatial multi-omics data. MultiSP learns a joint representation from multiple modalities and reconstructs denoised data, which facilitates the discovery of tissue spatial domains and the inference of spatially multimodal cell-cell communication with high biological fidelity.

Mo et al., 2026, Cell Genomics 6, 101141
May 13, 2026
https://doi.org/10.1016/j.xgen.2026.101141

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Technology

MultiSP deciphers tissue structure and multicellular communication from spatial multi-omics data

Chenfeng Mo,[1] Xiufen Zou,[1] and Suoqin Jin[1,2,*]
[1]School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China
[2]Lead contact
*Correspondence: sqjin@whu.edu.cn
https://doi.org/10.1016/j.xgen.2026.101141

SUMMARY

Recent breakthroughs in spatial multi-omics enable simultaneous profiling of different modalities while preserving tissue architecture, providing unprecedented opportunities to explore tissue complexity. However, due to the sparse and noisy nature of the data, interpreting these complex tissue structures and cellular communication remains challenging. We present MultiSP, a deep learning framework that enhances data representation through efficient spatial and feature similarity fusion, modality-specific probabilistic generative modeling, and cross-modality adversarial learning. Applied to various spatial multi-omics datasets, it outperforms existing methods in capturing biologically interpretable spatial domains. MultiSP also denoises spatial data, uncovers modality-specific spatial variations, and reveals gene regulation mechanisms. In the tumor microenvironment, it unravels fine-resolution cellular distribution maps, such as spatially neighboring macrophage-enriched sub-regions with distinct prognosis outcomes. Additionally, MultiSP facilitates the inference of spatially multimodal cell-cell communication. Together, MultiSP serves as a powerful framework for uncovering spatially multimodal heterogeneity and communication by integrating complementary information from multiple modalities.

INTRODUCTION

Spatial omics has recently progressed from single-modality techniques, such as spatial transcriptomics, to spatial multi-omics, which simultaneously captures multiple molecular layers—including gene expression, chromatin accessibility, and protein abundance—while preserving tissue architecture.[1] Emerging technologies, such as spatial protein and transcriptome sequencing (SPOTS),[2] microfluidic indexing-based spatial assay for transposase-accessible chromatin and RNA sequencing (MISAR-seq),[3] spatial assay for transposase-accessible chromatin using sequencing-RNA sequencing (spatial ATAC-RNA-seq),[4] deterministic barcoding in tissue for spatial omics sequencing (spatial DBiT-seq),[5] spatial co-indexing of transcriptomes and epitopes for multi-omics mapping by highly parallel sequencing (spatial CITE-seq),[6] and Slide-tags,[7] provide an unprecedented opportunity for investigating cellular spatial heterogeneity and communication.

A number of computational methods have been developed for integrating single-cell multi-omics data,[8] such as Seurat WNN,[9] totalVI,[10] MultiVI,[11] scCross,[12] and scAI,[13] as well as for identifying spatial domains from spatial transcriptomics,[14] such as SpaGCN,[15] STAGATE,[16] GraphST,[17] CellCharter,[18] stLearn,[19] and scNiche,[20] and from spatial chromatin accessibility data such as INSTINCT.[21] However, these methods either neglect spatial information or focus on a single omics modality, limiting their ability to achieve comprehensive characterization of spatially multimodal heterogeneity. Recently, algorithms specifically designed for integrating spatial multi-omics data, such as SpatialGlue,[22] spaVAE,[23] PRESENT,[24] COSMOS,[25] MISO,[26] stClinic,[27] and spaMGCN,[28] and integrating spatial transcriptomics with single-cell multi-omics data, such as SpaTrio[29] and SIMO,[30] have been proposed. These methods integrate multi-omics data through various fusion strategies and show improvements in downstream tasks such as spatial domain identification. These fusion strategies are broadly categorized into two types: weighted summation and feature concatenation. Weighted summation assigns modality-specific weights either by predefined criteria (e.g., COSMOS) or via learned attention mechanisms (e.g., SpatialGlue, CORAL,[31] and CANDIES[32]). By contrast, concatenation-based methods either directly combine the features (e.g., spaVAE and spaMGCN) or further integrate them using neural networks after concatenation (e.g., PRESENT and PRAGA[33]). Furthermore, MISO explicitly models inter-modal interactions during integration. Beyond fusion, some methods also perform multimodal alignment. For example, SpatialGlue introduces a correspondence loss, and PRESENT designs an inter-omics alignment loss, while SpaTrio and SIMO leverage optimal transport to align representation manifolds across modalities. Despite these advances, current approaches either overlook

the intrinsic distribution characteristics of different modalities, over-rely on spatial neighborhood structures, lack denoising capabilities, or fail to properly align modalities into a common low-dimensional space prior to fusion. Consequently, the effective integration of spatial multi-omics data and the deciphering of spatially multimodal heterogeneity remain significant challenges.

Beyond the characterization of tissue structure from different modalities, spatial multi-omics also provide the opportunity to better decipher how these spatial cellular structures communicate with each other in the tissue,[34,35] which has not been well explored in current spatial multi-omics data analysis. In addition, existing studies of cell-cell communication inference, including our previously developed CellChat method,[36,37] rely solely on mRNA expression as a proxy for cell-surface protein levels, assuming that mRNA abundance correlates well with protein expression. This assumption often fails, particularly for cell-surface proteins, as mRNA and protein expression are influenced by post-transcriptional regulation and the complex processes governing protein trafficking to the plasma membrane. On the other hand, secreted proteins are often difficult to stain for immunofluorescence.[38] Therefore, spatial multi-omics data bring new opportunities to analyze spatial cell-cell communication in native tissues by using expression profiles of cell-surface proteins.

To this end, we present MultiSP to reveal multicellular tissue structures and communication from spatial multi-omics data. MultiSP is a deep learning framework that incorporates both the neighborhood structure of cells within tissues and the distinct statistical distribution characteristics of each omics modality. Unlike other spatial multi-omics integration methods, MultiSP employs generative adversarial networks (GANs)[39] to align the variational autoencoder (VAE)[40]-derived representation manifolds of different omics modalities. While GANs have been utilized in single-cell omics for tasks such as batch correction and data generation,[41,42] their application for vertical data integration remains unexplored. To mitigate potential over-integration and preserve modality-specific information, we adopt a weighted adversarial alignment strategy. We evaluate MultiSP on datasets from various technologies and tissue types. Compared with other methods, MultiSP consistently achieves more accurate spatial domain detection and shows its powerful ability to denoise data and uncover spatial multimodal heterogeneity. Additionally, we perform an ablation study to evaluate the contributions of MultiSP's architectural components, highlighting the significance of each tailored component in driving its performance. Finally, we extend our CellChat framework by directly incorporating spatial locations of cells, gene expression of secreted ligands, and protein expression of corresponding surface receptors, enabling the inference of spatially multimodal cell-cell communication between interacting cell niches in native tissues.

DESIGN

MultiSP takes raw count matrices from multiple omics data and spatial location coordinates as inputs, employing a dedicated VAE for each omics modality (Figure 1A). After preprocessing the omics data, a spatial proximity graph (spatial graph) is constructed based on spatial coordinates, and a feature similarity graph (feature graph) is created using omics-specific features. The fused graphs based on the spatial graph and the feature graph, along with the preprocessed omics features, are fed into omics-specific encoders to generate low-dimensional latent embeddings for each cell. Finally, omics-specific decoders reconstruct the input data from these latent embeddings, according to the unique distribution characteristics of each omics type. To align and integrate the latent embeddings from different omics, MultiSP introduces a discriminator to distinguish between the latent embeddings of different omics. The VAE networks for each omics act as generators, aiming to produce latent embeddings that confuse the discriminator, making it difficult to differentiate between the omics representations. This adversarial interaction between the generators and the discriminator enables the alignment of latent manifolds from different omics data. To preserve meaningful modality-specific information, we adopt a weighted adversarial alignment strategy by introducing cell-specific weights in the adversarial loss. Multilayer perceptron (MLP) was used to fuse these low-dimensional embeddings, yielding a comprehensive and information-rich representation.

The resulting low-dimensional embeddings and reconstructed data generated by MultiSP can be applied to various downstream analyses, including spatial domain identification, data visualization, differential multimodal signatures analysis, and data denoising (Figure 1B). Moreover, based on our previously developed cell-cell communication inference method, CellChat, and the denoised data, we infer multimodal cell-cell communication between interacting spatial domains by calculating an interaction score based on the average gene expression values of a ligand and the average protein expression of a surface receptor (Figure 1B; STAR Methods). Given that most cell-cell communication events occur within a short spatial distance, we consider the biologically realistic interactions based on the maximal possible molecular interaction/diffusion ranges. For secreted signaling, we use the ideal diffusion range in a free medium, that is, the maximally allowable transport distance for small diffusive molecules (by default, 250 μm). For contact-dependent signaling, the interaction range is restricted to the nearest neighbors of each cell, such that signaling and target cells are in direct contact.

RESULTS

MultiSP facilitates accurate spatial domain detection in spatial RNA-ADT omics data

We first evaluated MultiSP on a human lymph node dataset generated using the CytAssist Visium platform (10× Genomics),[22] which simultaneously measures both RNA and ADT (antibody-derived tag) modalities. This dataset includes ground-truth labels obtained from the previous study,[22] where tissue domains were annotated by clinical experts based on hematoxylin and eosin (H&E)-stained sections (Figure 2A). We compared MultiSP with single-cell multi-omics integration methods such as Seurat and totalVI, spatial single-omics integration methods that only use gene expression data, such as STAGATE and GraphST, as well as spatial multi-omics integration methods, such as SpatialGlue, PRESENT, and MISO. Additionally, we examined the benefits of utilizing two modalities by

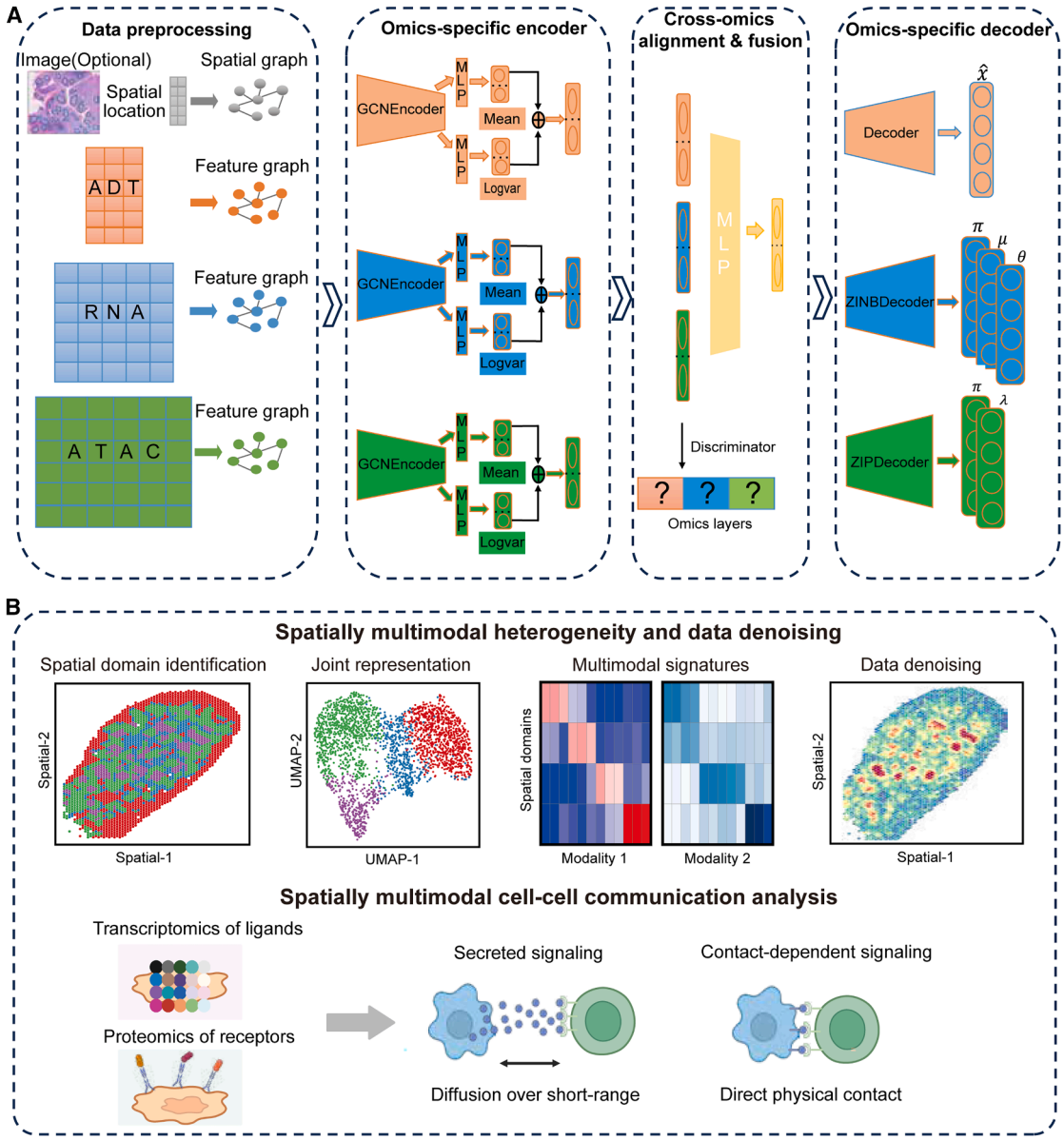


Figure 1. Overview of MultiSP

(A) Inputs include RNA, ATAC, ADT, spatial coordinates, and an optional H&E-stained tissue image. These are transformed into two complementary graphs: a spatial neighborhood graph and a feature similarity graph. Within a VAE framework, each omics modality is processed by a dedicated encoder-decoder pair. Each encoder employs a GCN module followed by multi-layer perceptrons (MLPs) to generate a latent representation. A discriminator aligns the latent spaces across modalities, and a final MLP integrates them into a unified embedding. The decoders are modality specific, each designed to match the data distribution of its respective omics.

(B) MultiSP supports multiple downstream analyses: spatial domain identification, joint visualization, differential multimodal signature detection, and data denoising. Denoised spatial transcriptomics (for secreted ligands) and proteomics (for surface receptors) can be integrated to infer spatially resolved multimodal cell-cell communication, distinguishing between secreted-based and contact-dependent signaling mechanisms. The schematic diagram of cell-cell communication was created in BioRender (https://BioRender.com/kr4y5em).

including the results that only use single-omics data, labeled as RNA and Protein, respectively. We first visually examined the spatial domains identified by each method (Figure 2B). As shown in Figures 2A and 2B, most methods successfully detected large-scale domains such as the cortex (ctx), pericapsular adipose tissue, and follicle. However, the spatial domains identified by MultiSP aligned more closely with the ground truth. totalVI, STAGATE, GraphST, and MISO failed to detect the small domain hilum, while Seurat, SpatialGlue, and PRESENT tended to merge the hilum with the surrounding pericapsular adipose tissue. In contrast, MultiSP accurately identified the hilum and clearly distinguished it from the adjacent pericapsular adipose tissue.

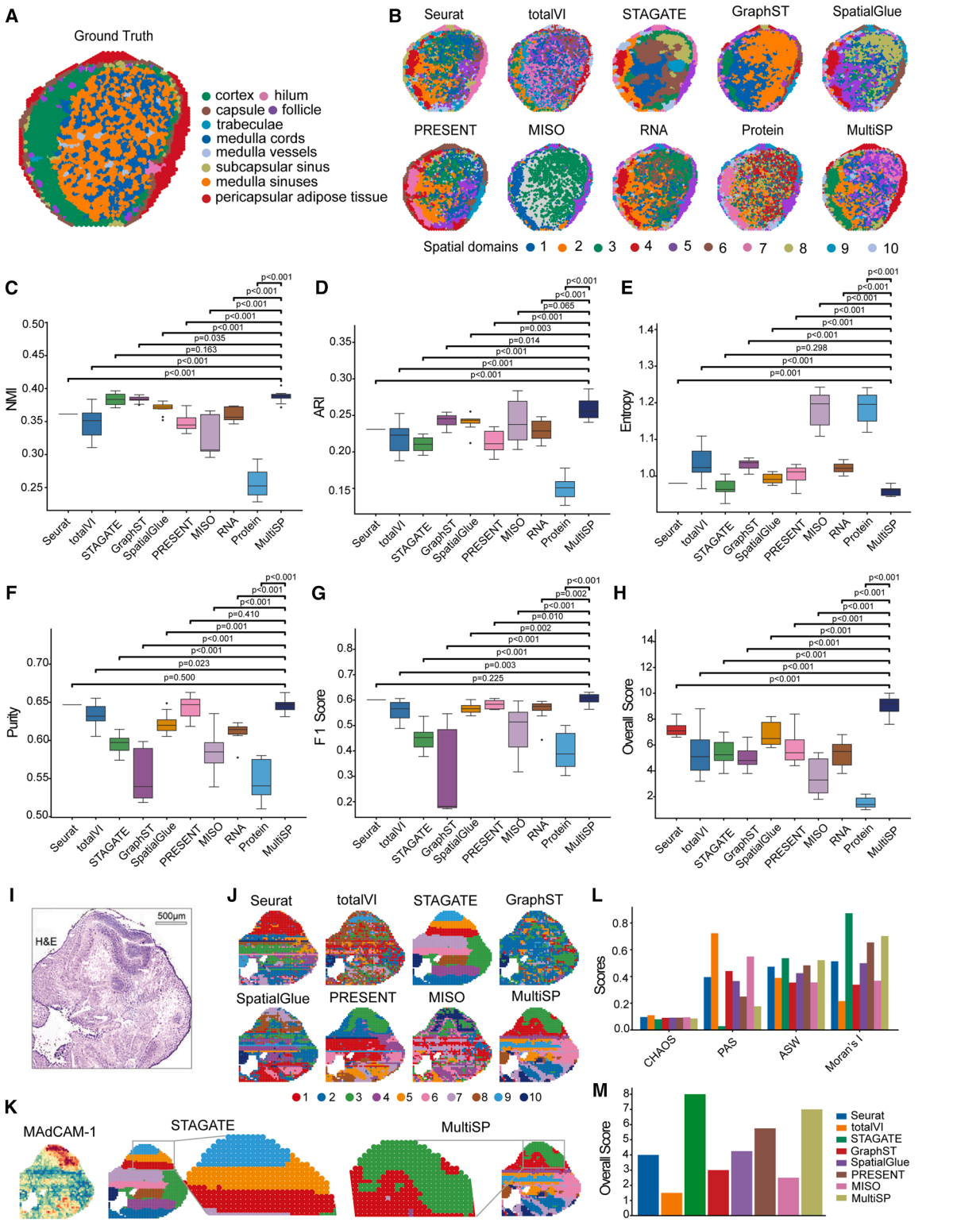


Figure 2. MultiSP facilitates accurate spatial domain detection in spatial RNA-ADT omics data

(A) Annotated ground-truth domain labels of human lymph node dataset.

(B) Spatial domains identified by all methods as well as MultiSP under single-modality analysis for RNA or protein.

(legend continued on next page)

Furthermore, we employed five supervised metrics to quantify the accuracy of spatial domain detection, including NMI (normalized mutual information), ARI (adjusted Rand index), entropy, purity, and F1 score. We also defined an overall accuracy score as the average rank across all five metrics. Better performance corresponds to higher values for NMI, ARI, purity, F1 score, and overall score and lower values for entropy (STAR Methods). MultiSP significantly outperforms all other methods on these metrics, with only one non-significant exception versus a comparator method on a certain metric. As expected, MultiSP achieves a significantly higher overall score than all other methods (Figures 2C–2H). Among single-modality results, either RNA-only or protein-only analyses cannot capture the spatial domains well, particularly noticeable in the protein-only results, suggesting the importance of integrating different modalities for revealing spatial heterogeneity.

We then applied MultiSP to a mouse embryo dataset generated using the spatial DBiT-seq technology to demonstrate its effectiveness and robustness in integrating transcriptomic and proteomic data. Notably, only MultiSP aligned more closely with the previously reported MAdCAM1+ forebrain region that also correlated with tissue morphology of the mouse embryo brain[5] (Figures 2I–2K). PRESENT also showed good performance, but SpatialGlue and MISO only captured a portion of this region. Other methods, including Seurat, totalVI, STAGATE, and GraphST, showed worse performance in capturing this biologically meaningful spatial domain. Moreover, we evaluated different methods using four unsupervised metrics (i.e., CHAOS, PAS, ASW, and Moran's I) and an overall score that integrates these four metrics (STAR Methods). MultiSP and STAGATE exhibited better performance over other methods in terms of these four metrics (Figures 2L and 2M). While STAGATE shows a higher overall score than MultiSP, STAGATE did not correctly capture this MAdCAM1+ forebrain region (Figure 2K). This over-smoothed phenomenon of STAGATE was also observed in the previous study of SpatialGlue.[22]

In summary, MultiSP effectively integrates complementary spatial transcriptomics and proteomics data, enabling more accurate spatial domain detection.

MultiSP resolves mouse brain structures at different development stages from spatial RNA-ATAC omics data

Next, we assess the ability of MultiSP to integrate spatial transcriptomics and epigenomics data by applying it to the MISAR-seq mouse brain dataset,[3] which simultaneously measures transcriptomics and epigenomics across eight slices from four developmental stages: embryonic day (E)11.0, E13.5, E15.5, and E18.5, with two slices per stage. We used the four section 1 (S1) slices with anatomical annotations provided in the original study,[3] which served as the ground truth. Compared to the spatial transcriptomics and proteomics data used in the aforementioned section, this dataset exhibits higher dimensionality and greater sparsity. In comparison to other methods, including Seurat, MultiVI, STAGATE, GraphST, INSTINCT, SpatialGlue, PRESENT, MISO, and single-modality analyses labeled as RNA and ATAC, MultiSP achieves the best median score across all five supervised metrics and a significantly higher overall score when integrating the five metrics (Figures 3A–3F). This demonstrates that MultiSP can also achieve more accurate spatial domain detection when integrating spatial transcriptomics and epigenomics data.

Using the E15.5_S1 mouse brain sample as an example, we visually examined the identified spatial architecture from different methods, with 12 distinct domains based on the original annotations (Figure 3G). MISO demonstrated the weakest performance, failing to detect major spatial domains, leading to the worst performance in terms of the quantitative metrics (Figures 3A–3F). In contrast, other methods successfully identified key domains such as Dpallm (mantle zone of dorsal pallium), Dpallv (ventricular zone of dorsal pallium), subpallium, thalamus, and cartilage. Notably, only MultiSP, INSTINCT, and SpatialGlue accurately detected the cerebellar vermis and distinguished it from the midbrain. Compared to all other methods, the muscle identified by MultiSP exhibited superior spatial continuity and greater consistency with the ground truth. Raw gene expression and chromatin accessibility data are often corrupted by high levels of noise and high dropout events. By modeling the gene expression profile using a zero-inflated negative binomial distribution (ZINB) and the chromatin accessibility using a zero-inflated Poisson distribution (ZIP), MultiSP effectively denoises the original data, enabling a clearer representation of the spatial patterns of genes. We further performed differential gene expression and chromatin accessibility analyses based on the detected spatial domains and visualized the spatial distribution of gene and peak markers for specific domains before and after denoising (Figure 3H). Specifically, we highlighted six domains: clusters 1, 2, 9, 4, 6, and 12, corresponding to cartilage, Dpallm, muscle, cerebellar vermis, thalamus, and subpallium in the original annotations. The associated gene markers *Col1a1*, *Nfib*, and *Dlk1* and the marker peaks chr9:91350781–91351282, chr5:131596822–131597323, and chr2:158556614-158557115 were exclusively enriched in these domains, respectively. Notably, previous studies have shown that type I collagen, such as *Col1a1*, is highly enriched, reflecting its role in collagen synthesis and extracellular matrix (ECM) organization.[43] *Nfib* has been identified as a key regulator of neural stem cell differentiation and the axonal projection of corticofugal neurons,[44] and *Dlk1* is recognized as an important regulator of muscle development and differentiation.[45] Using the chromVAR package,[46] we performed enrichment analysis of transcription factor motifs based on chromatin accessibility and identified the motifs CREB1,[47] MEIS2,[48] and DLX6[49] associated with the indicated specific marker peaks. These motifs were also biologically associated with the cerebellar vermis, thalamus, and subpallium,

(C–H) The quantitative comparison of different methods using five evaluation metrics (C–G), as well as the overall score (H) (n = 10 independent experiments). p values were calculated using a one-sided unpaired Wilcoxon rank sum test.

(I–M) Application to spatial DBiT-seq mouse embryo dataset. (I) H&E-stained histology image. (J) Spatial domains identified by all methods. (K) Magnified clustering results of the spatial region marked by the protein MAdCAM-1. (L and M) The quantitative comparison of different methods using four unsupervised evaluation metrics and the overall score.

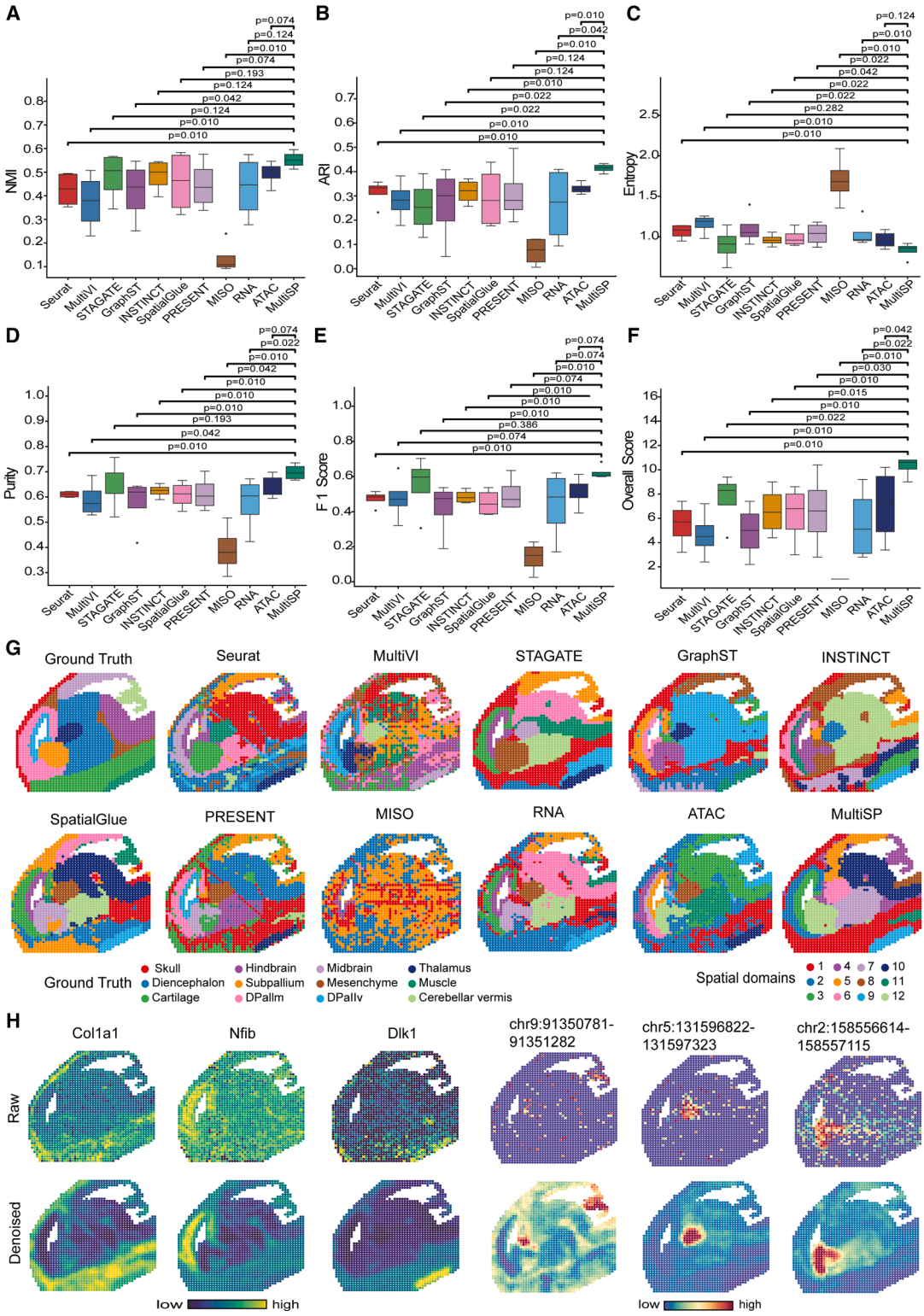


Figure 3. MultiSP resolves mouse brain structures at different development stages from spatial RNA-ATAC omics data

(A–F) The quantitative comparison of different methods on the MISAR-seq dataset. Each boxplot contains results from samples across four tissue slices: E11_0-S1, E13_5-S1, E15_5-S1, and E18_5-S1. *p* values were calculated using a one-sided unpaired Wilcoxon rank sum test.

(legend continued on next page)

respectively. Together, MultiSP is able to identify the biologically meaningful marker genes and peaks that are consistent with the known biology of anatomical annotations. Compared to the raw data that contained significant noise, the denoised data produced by MultiSP exhibited more enriched and well-defined spatial patterns. The results with all other slices are shown in Figure S1.

These results demonstrate that MultiSP effectively integrates complementary information from spatial transcriptomics and epigenomics, enabling the detailed dissection of complex spatial tissue structures.

MultiSP accurately captures mouse brain anatomy, enables spatial data denoising, and reveals spatial patterns of gene regulations by integrating spatial epigenomics and transcriptomics

To demonstrate the ability of MultiSP to denoise spatial data and unravel spatial regulatory mechanisms, we applied MultiSP to a postnatal day (P)22 mouse brain coronal section dataset generated using spatial ATAC-RNA-seq technology, which simultaneously measures mRNA and open chromatin regions. Using the Allen Brain Atlas[50] as a reference to annotate anatomical domains (Figure 4A), we tested MultiSP against other methods for benchmarking (Figure 4B). While both RNA and ATAC modalities could identify the vl (lateral ventricle), the RNA modality more clearly distinguished the ccg (genu of corpus callosum) but failed to differentiate between the different layers of the ctx (cortex layers). In contrast, the ATAC modality was able to identify the small region lpo (lateral preoptic area) and distinct ctx layers, which may explain why spatial transcriptomics methods such as STAGATE and GraphST are unable to achieve this. MultiSP outperformed other methods by accurately identifying domains in the reference annotation, including the cp (caudoputamen), acb (nucleus accumbens), ccg, ls (lateral septal nucleus), vl, and lpo, and providing a more refined segmentation of the ctx. The atlas annotation demonstrated a pronounced spatial pattern in the dataset, which non-spatial, single-cell multi-omics methods, such as Seurat and MultiVI, and the spatial multi-omics methods PRESENT and MISO failed to fully capture, often producing substantial noise.

Although SpatialGlue performed comparatively well, with the ability to identify most domains, MultiSP generated smoother and more continuous spatial domains. This observation is further supported by the unsupervised quantitative metrics of spatial continuity (Figures 4C and 4D). These results suggest the superior and robust performance of MultiSP in capturing mouse brain anatomy by integrating gene expression and chromatin accessibility profiles.

Next, we visualized the spatial distribution of marker genes across six distinct domains before and after denoising (Figure 4E). The gene markers *Rreb1*,[51] *Dgkg*,[52] *Mbp*,[53] *Sox4*,[54] *Pde10a*,[55] and *Snap25*[56] correspond to the lpo, ls, ccg, vl, cp, and ctx domains, respectively. Our analysis revealed that the original gene expression data contained significant noise, whereas the denoised gene expression produced by MultiSP exhibited enriched and well-defined spatial patterns. We further quantitatively assess the denoising performance by employing the widely used strategy in the previous benchmarking study,[57] that is, the evaluation of the clustering performance on denoised data. We compared MultiSP with established denoising methods, including Sprod[58] and spaVAE.[23] By evaluating the clustering performance on the denoised spatial gene data, we observed superior denoising performance of MultiSP compared to the raw data, Sprod, and spaVAE (Figures 4F, 4G, S2A, and S2B). These findings demonstrate that MultiSP effectively integrates spatial and multimodal information to denoise gene expression, facilitating its application in various downstream analyses.

Complex regulatory mechanisms exist between gene expression and chromatin accessibility.[59] By integrating these two modalities in a spatial manner, we are able to explore the spatial patterns of gene regulation. Transcription factors play a critical role in cell differentiation and influence cell fate decisions during brain aging and development by regulating gene expression.[60–62] Using the chromVAR package,[46] we performed enrichment analysis of transcription factor motifs based on chromatin accessibility data. Combined with denoised gene expression data, we identified distinct gene regulatory relationships for the marker genes *Neurod2* and *Olig1* in the mouse brain (Figure 4H). Specifically, the regulatory factor Neurod2 exhibited synchronous changes in gene and motif activity across different domains, with a similar spatial pattern of gene expression and motif activity peaking at the ctx. In contrast, the regulatory factor Olig1 showed a markedly different trend and spatial distribution. We conducted a spatial colocalization analysis of the two motif-gene pairs using the Lee statistical test.[63] Expectedly, we observed a colocalized pattern between motif MA0668.1 and gene *Neurod2* and a non-colocalized pattern between motif MA0826.1 and *Olig1*, confirming their distinct regulatory patterns (Figure S2C).

MultiSP uncovers spatially neighboring tumor-associated macrophage-enriched regions with distinct prognosis power

To analyze the highly variable and complex tumor microenvironment,[64] we applied MultiSP to a mouse breast cancer dataset generated using the SPOTS technology, which simultaneously measures spatial transcriptomics and proteomics features. By jointly analyzing the two modalities, MultiSP identified seven spatial domains, which were spatially organized in the tissue and well separated in the learned uniform manifold approximation and projection (UMAP) space (Figure 5A). These domains were labeled as peritumoral, plasma cells, two macrophage-enriched regions (Mac-1 and Mac-2), two fibroblast-enriched regions (Fib-1 and Fib-2), and epithelial cells, respectively. Next, we conducted differential gene and protein expression analyses to identify markers for each domain relative to others, and we

(G) The spatial visualization colored by ground truth and spatial clusters identified based on different methods in the E15.5_S1 mouse brain sample. The spatial domains produced by MultiSP are numbered and indicated.
(H) Visualizations of the raw and MultiSP denoised values of the representative marker genes and peaks.

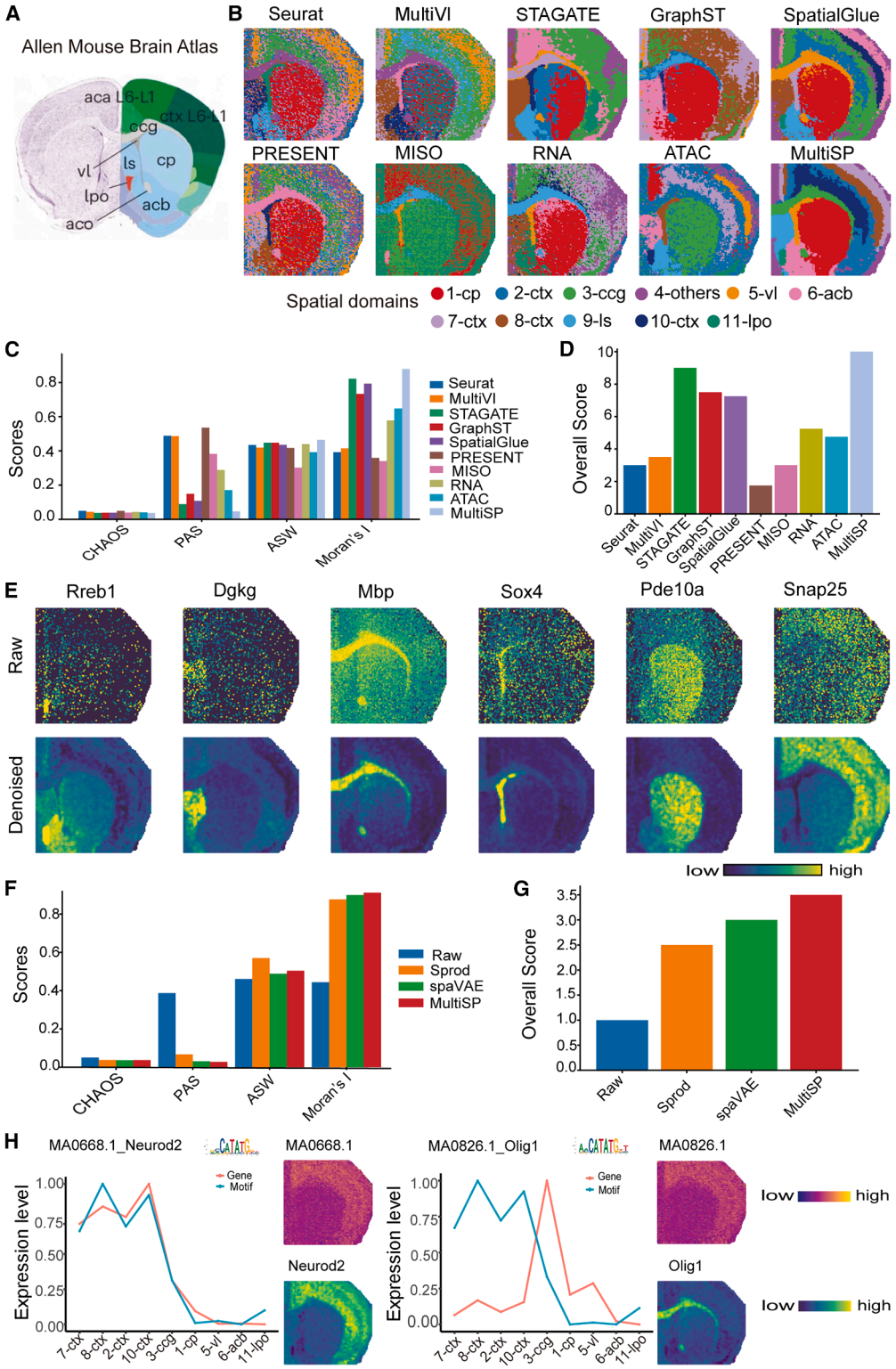


Figure 4. MultiSP captures mouse brain anatomy from spatial ATAC-RNA-seq data and enables spatial data denoising

(A) Annotated reference from the Allen Mouse Brain Atlas.

(B) Spatial domains identified by different methods. The spatial domains produced by MultiSP are numbered and indicated.

(C and D) The quantitative comparison of different methods.

(legend continued on next page)

presented the spatial distributions of the representative markers (Figures 5B and 5C). Overall, we found that gene markers exhibited more pronounced differential expression patterns than protein markers. Notably, the gene marker *Krt8* and the protein marker CD326 were both enriched in the epithelial cells, and the gene marker *Hba-a2* and the protein marker CD19 effectively distinguished the peritumoral area from the macrophage-enriched region. Additionally, the gene markers *Trim33* and *Igha* and the protein marker podoplanin (PDPN) better reflected the differences among Mac-2- and plasma cell-enriched regions. Previous study shows that PDPN is upregulated in cancer-associated fibroblasts and inflammatory macrophages that contribute to cancer progression,[65] suggesting that targeting Mac-2-enriched regions with relatively high PDPN may inhibit cancer progression.

Intriguingly, MultiSP deciphered fine-resolution cellular distribution maps of fibroblasts and macrophages, including two fibroblast-enriched sub-regions (Fib-1 and Fib-2) and two macrophage-enriched sub-regions (Mac-1 and Mac-2). In addition to the aforementioned marker genes and proteins that discriminated these different sub-regions, we further systematically characterized their differences by performing gene scoring analysis using well-established gene sets (STAR Methods). We found that significantly higher gene scores of ECM organization and COLLAGEN formation were observed in Fib-2 compared to Fib-1 (Figure 5D), suggesting that Fib-2 produces an excessive ECM, forming a barrier that modulates cancer invasion and immune response. This observation was consistent with the fact that Fib-2, rather than Fib-1, was spatially close to the peritumoral region (Figure 5A). Moreover, the two fibroblast subsets exhibited distinct gene expression patterns (Figure 5D). Fib-1 showed high expression of genes involved in the regulation of the immune process, such as *Tyrobp*, *Pycard*, *Arg1*, *Xbp1*, and *Appl2*, which was consistent with the important roles of fibroblasts in regulating immune responses.[45,66] Fib-2 represents a fibroblast state that promotes cancer progression through ECM remodeling, where genes involved in ECM organization, such as *Lgals3*, *Hspg2*, *Fn1*, *Fap*, and *Mmp14*, were enriched in Fib-2.[45,66] For the two macrophage-enriched sub-regions, we observed the highest gene scores of proliferation and wound healing in Mac-2 and the lowest values in Mac-1. Using the signature genes of M2-like macrophages, we found significantly higher gene scores in Mac-2-enriched regions compared to Mac-1 (Figure 5E). Furthermore, we conducted a differential gene expression analysis between Mac-1- and Mac-2-enriched regions (Figure 5F). Consistent with the gene scoring analysis, the Mac-1-enriched region showed high expression of inflammation-associated factors such as *Aqp5* and *Xbp1*,[67] while the Mac-2-enriched region was characterized by alternatively activated macrophage marker genes, including *Mmp2* and *Spp1*,[68–70] as well as genes related to the matrix and structure organization, such as *Col4a1* and *Vim*.

Moreover, we asked whether the spatially neighboring Mac-1- and Mac-2-enriched regions exhibited distinct prognostic power by performing survival analysis. Since protein abundance cannot well distinguish different regions (Figure 5C), we performed this analysis using gene expression data. By utilizing the bulk RNA-seq data of The Cancer Genome Atlas (TCGA) breast cancer patients from UCSC Xena[71] and using the top 50 differentially expressed genes (DEGs) between Mac-1- and Mac-2-enriched regions, we performed survival analysis using the survival R package (STAR Methods). GSVA (gene set variation analysis) scoring was used to score each patient based on the expression of DEGs. Significant differences in the prognostic abilities of Mac-1- and Mac-2-enriched regions were observed (Figure 5G). When using marker genes of Mac-1, the patients with high GSVA scores exhibited higher survival rates compared to the remaining patients ($p = 0.047$, two-sided log rank test). On the contrary, the group of patients with high GSVA scores of marker genes of the Mac-2-enriched region had significantly poorer prognosis compared to the remaining group of patients ($p = 0.021$). By using varying numbers of DEGs identified in the Mac-2-enriched region, we consistently observed poor prognosis for this region (Figure S3). In addition, we observed the consistency between the identified DEGs (Figure 5F) and known poor-prognosis macrophage signatures in human breast cancer, such as *IGFBP2*,[72] *PDPN*,[73] *AGRN*,[74] and *SPP1*,[75] suggesting the clinical relevance of our findings. These results indicate that the two identified macrophage-enriched regions are biologically distinct within the complex tumor microenvironment. Together, by jointly analyzing spatial transcriptomics and proteomics, MultiSP depicted refined spatial structures and identified a macrophage-enriched region associated with poor prognosis, promoting tumor growth by creating an immunosuppressive microenvironment and barrier at tumor borders.

Additional applications to a broader range of spatial multi-omics datasets from various technologies and tissues

To showcase MultiSP's broad applicability to various tissues and data sources, we applied it to a wide range of spatial multi-omics datasets derived from various technologies and tissue types. These include E13 mouse embryo data from spatial ATAC-RNA-seq (RNA+ATAC),[4] human skin data from spatial CITE-seq (RNA+ADT),[6] and human tonsil dataset from 10× Genomics (RNA+ADT+Image).[76]

For the spatial ATAC-RNA-seq mouse embryo dataset (Figure S4A), MISO failed to detect the eye region. Seurat, MultiVI, STAGATE, GraphST, and PRESENT identified the eye with partial overlap with other domains. Only MultiSP and SpatialGlue successfully detected the complete eye region and clearly separated it from other anatomical structures (Figure S4B). We further evaluated the performance of each method using the four unsupervised metrics. Among them, MultiSP and STAGATE consistently outperformed other

(E) Visualizations of the raw gene expressions and MultiSP denoised values of the six representative marker genes.
(F and G) Clustering performance of raw and denoised spatial gene expression data from various methods.
(H) The scaled denoised gene expression and ChromVAR-derived motif activity of the two distinct regulators across different domains from the outer layer to the inner layer.

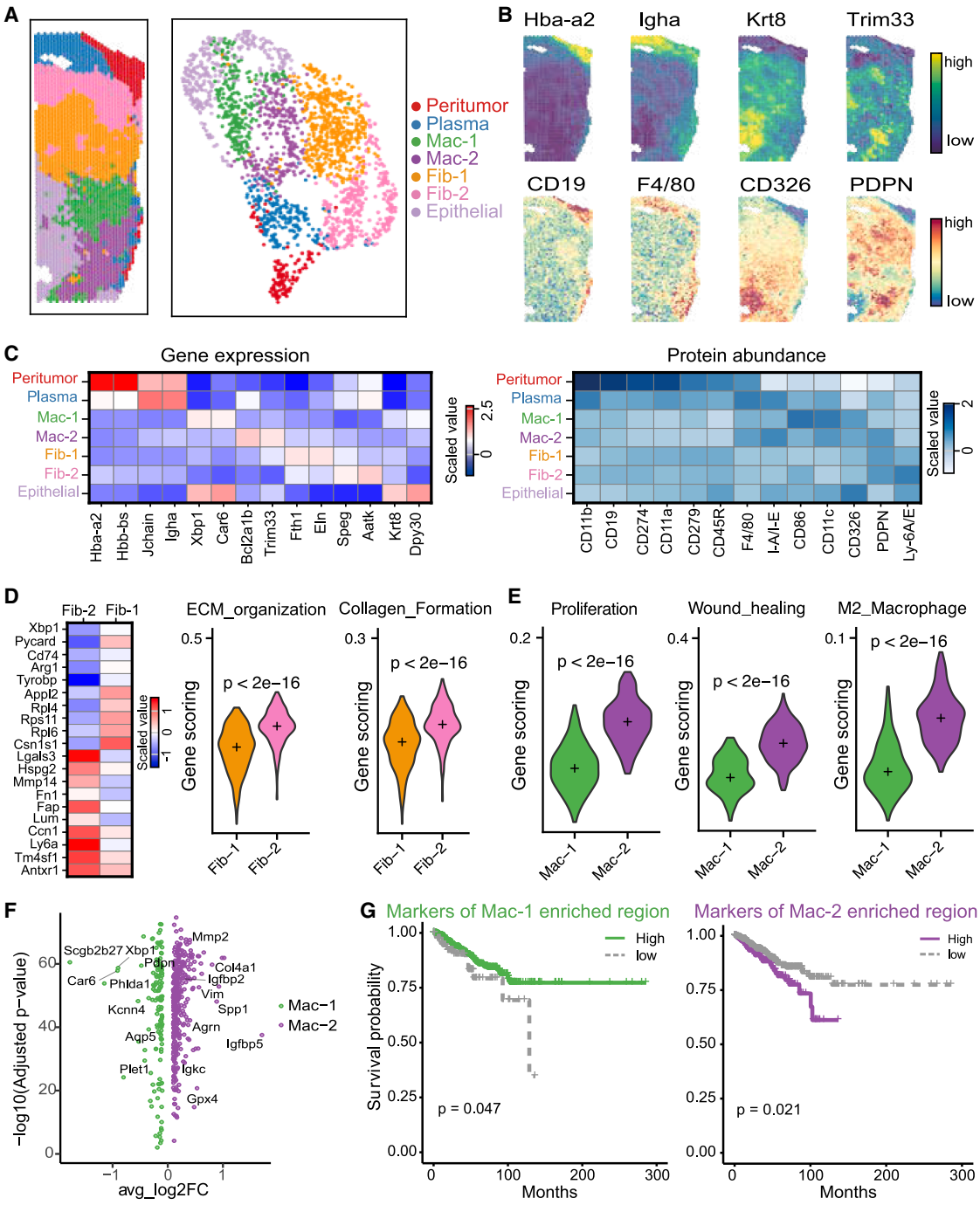


Figure 5. MultiSP uncovers fine-resolution spatial regions with distinct prognostic power

(A) The spatial and UMAP visualization of clusters identified based on MultiSP in the mouse breast cancer dataset.

(B) Normalized RNA levels of *Hba-a2*, *Igha*, *Krt8,* and *Trim33* and normalized ADT levels of CD19, F4/80, CD326, and PDPN.

(C) The relative denoised averaged RNA expression levels and ADT levels across clusters.

(D) The differential expressed genes between Fib-1 and Fib-2 (left) and the violin plot of calculated gene scores for given gene sets among the two fibroblasts enriched regions (right). In the violin plot, Fib-1- and Fib-2-enriched spots were used for gene score calculation. *p* values are calculated from a two-sided Wilcoxon rank sum test.

(legend continued on next page)

methods across all four metrics. Although STAGATE achieved a higher overall score than MultiSP, it failed to accurately distinguish between the stratum (pallium) and stratum (subpallium) domains, which display distinct morphologies in the H&E images (Figures S4A and S4C–S4F).

For the spatial CITE-seq human skin dataset, MultiSP, PRESENT, and Seurat accurately captured the dermis region, whereas other methods failed to do so. Among all methods, MultiSP identified the pilosebaceous unit with the greatest continuity and the closest alignment to the corresponding histology image (Figures S5A and S5B). In terms of the four unsupervised metrics, MultiSP and PRESENT exhibited better performance than the other methods (Figure S5C).

Given the flexibility of the MultiSP framework, it can also handle three or more modalities. For instance, H&E-stained images can be treated as an additional modality. We integrated preprocessed image features into the MultiSP architecture and reconstructed them using a standard mean squared error (MSE) decoder. We demonstrated the performance of MultiSP using the human tonsil dataset from 10× Genomics, which included spatial gene expression, protein, and histology image data (Figures S6A and S6B). Specifically, we followed the original study of MISO to assess each method's ability to localize known germinal centers and calculated the F1 score to evaluate the accuracy. MultiSP achieved the highest median F1 score compared to other methods, including MISO, SpatialGlue, PRESENT, and stLearn, as well as the single-omics methods Seurat and totalVI (Figure S6C). Together, these comprehensive comparisons demonstrate the flexibility and generalizability of the MultiSP framework.

Harnessing spatial multi-omics data to refine the reliability of cell-cell communication events

The spatial multi-omics data offer an unprecedented opportunity to analyze cell-cell communication by utilizing the mRNA expression of a ligand and the ADT expression of a surface receptor. We extended our previous cell-cell communication inference method CellChat to infer spatially multimodal cell-cell communication between interacting cell groups (Figure 6A; STAR Methods).

Taking the mouse spleen data that simultaneously measure spatial transcriptomics and proteomics using the SPOTS technology[2] as an example, we first applied MultiSP to identify the spatial domains. The spleen, a vital lymphoid organ essential for immune function, plays a key role in hematopoiesis and the clearance of aging red blood cells.[77] MultiSP identified four distinct spatial domains, which showed spatially compartmentalized structures in the tissue that were well separated in the learned UMAP space (Figure 6B). These domains were labeled as two macrophage subset-enriched regions (Mac-1 and Mac-2), along with a B cell-enriched region and a T cell-enriched region based on the enrichment of known marker genes. Of note, these spatial domains cannot be considered pure cell types due to the low-resolution spatial technology. We performed differential gene and protein expression analyses for each spatial domain (Figure 6C) and identified protein markers distinguishing B cells (e.g., CD19, CD169, and B220), T cells (e.g., CD3, CD4, and CD8), and macrophages (e.g., F4_80, CD163, and CD68), which well aligned with the findings from the original study.[2] However, protein markers alone were insufficient to differentiate between the two macrophage subtypes. The differences between Mac-1 and Mac-2 were primarily revealed by gene markers, with Mac-1 characterized by genes such as *Hbb-a1*, *Car2*, and *Pf4* and Mac-2 enriched with *Igkc*, *Igha*, and *Ighm*. We also visualized the spatial expression distribution of T cell protein markers (Cd4), key macrophage protein markers (F4_80), and gene markers for the two macrophage subsets (*Pf4* for Mac-1 and *Igkc* for Mac-2) (Figure 6D), revealing distinct expression patterns in their respective domains. These results highlight MultiSP's ability to capture the unique contributions of each modality for more refined spatial domain detection.

Next, we analyzed cell-cell communication between the identified spatial domains and compared communication inferred from protein expression of cell-surface receptors with that based on the traditional mRNA expression approach for both ligands and receptors. To ensure fair comparison, we compared the two approaches on the same collection of candidate L-R (ligand-receptor) pairs, where the L-R pairs with receptors measured in ADT were used. We found that both mRNA and ADT levels of receptors in the inferred cell-cell communication using denoised RNA+ADT data were higher than those using the denoised RNA-only data and the raw RNA+ADT data (Figure 6E). In addition, we compared the methods that only used the RNA data, including COMMOT[78] and SpatialDM.[79] Consistent with the results from CellChat, which only used RNA data, we observed similar results for COMMOT and SpatialDM (Figures 6F and 6G), where worse performance was observed compared to the results using RNA+ADT data. Among the identified ligand-receptor interactions, when using RNA-only data, both T cell- and B cell-enriched regions respond to IL-16-CD4 signaling; however, when using RNA+ADT data, only the T cell-enriched region was the predominant target of IL-16-CD4 signaling (Figure 6H), which was consistent with the exclusively enriched CD4 protein expression in the T cell-enriched region (Figure 6D). Importantly, a previous study stated that *IL16* was a well-known modulator of T cell activation.[80] Similar results were observed for major histocompatibility complex (MHC) class I signaling mediated by the *H2-AB1* ligand and its receptor CD4, where the T cell-enriched region with elevated protein expression of CD4 responds to this cell-contact-dependent signaling (Figure 6I). In addition, we calculated the overlap in cell-cell communication under independent runs of MultiSP and observed the same enriched signaling, suggesting the

(E) The violin plot of calculated gene scores for given gene sets among the two macrophage-enriched regions. In the violin plot, Mac-1- and Mac-2-enriched spots were used for gene score calculation. *p* values are calculated from a two-sided Wilcoxon rank sum test.

(F) The differentially expressed genes between the two macrophage subsets.

(G) Overall disease-free survival curves of patient groups with high and low gene scores using the marker genes of Mac-1-enriched region (left) and Mac-2-enriched region (right), respectively. *p* values are calculated from a two-sided log rank test.

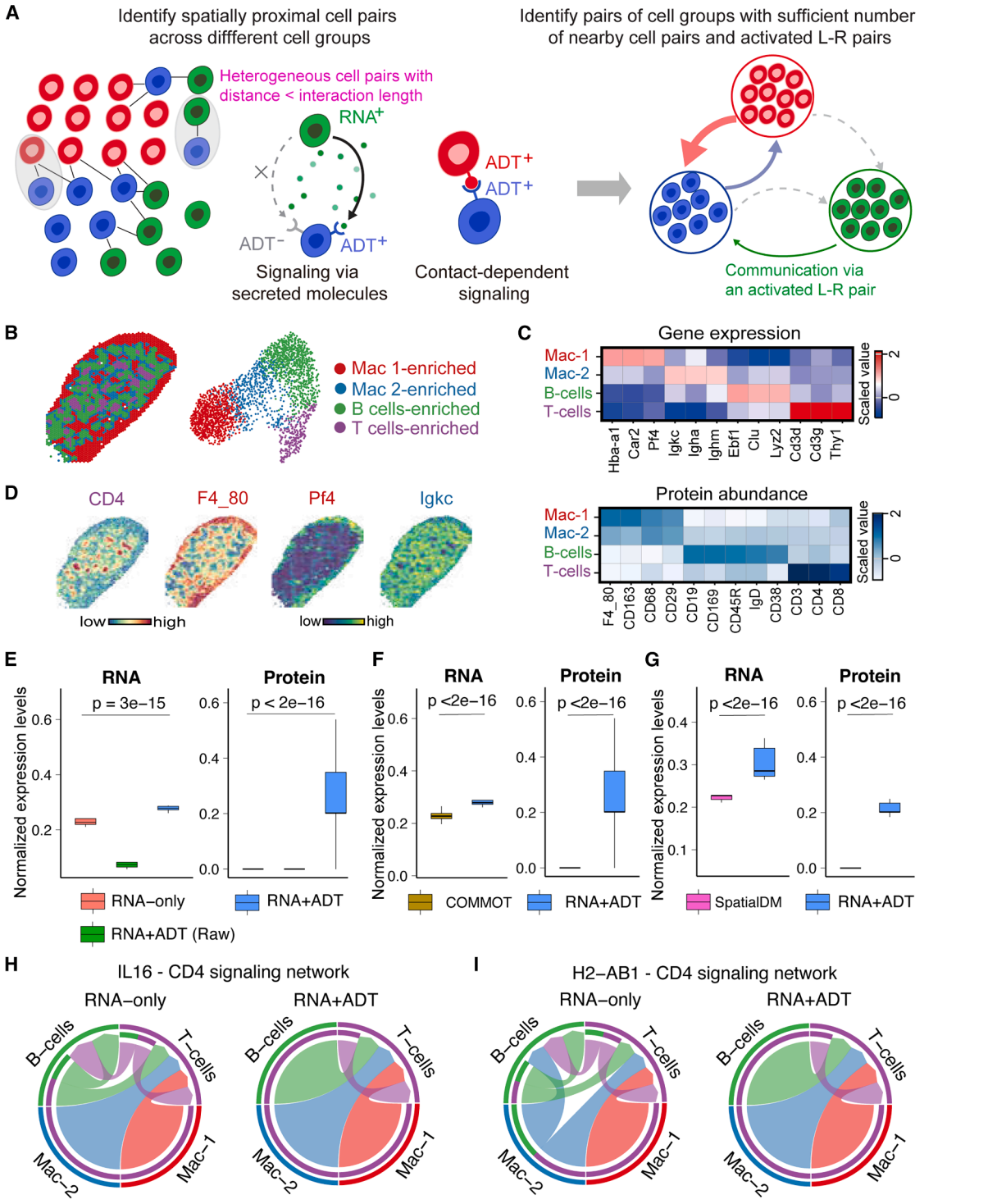


Figure 6. Harnessing both mRNA and ADT expression to detect reliable cell-cell communication events

(A) Spatially proximal multimodal cell-cell communication between adjacent groups of cells is inferred by first identifying pairs of cells that are physically proximal within the maximum assumed interaction/diffusion range and then identifying pairs of cell groups that have a sufficient number of nearby cell pairs.

(B) The UMAP and spatial visualization of domains identified based on MultiSP in the mouse spleen dataset.

(C) The average expression of each feature within each spatial domain and the difference of ADT expression levels (bottom) and RNA expression levels (top).

(D) Normalized ADT level of CD4 for T cells and F4_80 for macrophages and normalized RNA expressions of *Pf4* and *Igkc* for the two macrophage-enriched regions.

(legend continued on next page)

robustness of our method (Figure S7A). Collectively, our results suggest the importance of harnessing both mRNA and ADT expression in detecting reliable cell-cell communication events.

DISCUSSION

MultiSP is an efficient deep learning framework designed to learn latent representations of cells that integrate both spatial information and multimodal data, facilitating accurate spatial domain detection and subsequent downstream analyses such as low-dimensional visualization, differential multimodal signatures analysis, data denoising, and multimodal cell-cell communication. We demonstrated MultiSP's superior and robust performance on datasets from a variety of platforms, including 10× Genomics, SPOTS, MISAR-seq, spatial CITE-seq, spatial DBiT-seq, and spatial ATAC-RNA-seq.

Firstly, MultiSP outperformed competing methods in accurately identifying biological spatial domains in highly heterogeneous tissue samples with diverse omics combinations. Secondly, we demonstrated MultiSP's ability to denoise spatial omics data. Compared to the original data, the denoised data exhibited clear spatial patterns with enhanced spatial signaling. We also quantitatively assessed the denoising performance of MultiSP and compared it against two spatial denoising methods (i.e., Sprod and spaVAE) by evaluating clustering performance on denoised data from the P22 mouse brain (RNA modality), the MISAR-seq E15.5 mouse brain (ATAC modality), and the human lymph node (ADT modality) datasets. In all modalities, MultiSP consistently outperformed the raw data as well as the other two methods (Figure S8). To demonstrate the effectiveness of ZINB modeling in denoising gene expression data, we evaluated clustering performance on denoised data after removing ZINB and observed a noticeable decrease in spatial continuity and accuracy of the detected spatial domains (Figure S9). Finally, when applied to the tumor microenvironment in breast cancer tissue, MultiSP identified fine-resolution cellular distribution maps of fibroblasts and macrophages, including spatially adjacent tumor-associated macrophage-enriched regions that showed distinct prognostic power. We also identified several literature-supported macrophage signatures in human breast cancer, such as *IGFBP2*, *PDPN*, *AGRN*, and *SPP1*, suggesting the clinical relevance of the Mac-2-enriched-region-associated gene signatures. However, these findings would benefit from further experimental validation. Together, these results highlight the strong potential of MultiSP in analyzing the ever-growing spatial multi-omics data.

The successful performance of MultiSP lies in utilizing a probabilistic generative model that leverages cellular neighborhood similarity and modality-specific data distribution as well as an adversarial learning component to enhance data representation and cross-modality integration. By incorporating weighted adversarial alignment and a pretraining strategy, MultiSP not only projects different modalities into a joint latent space but also enables seamless integration of multiple modalities. Although MultiSP and SpatialGlue both construct spatial and feature graphs, MultiSP differs in several key aspects. First, it employs a VAE to refine the latent representation of each omics modality. Second, it concatenates latent embeddings before aligning modalities adversarially, allowing flexible integration of any number of omics types—unlike SpatialGlue, which aligns and fuses modalities pairwise. Third, MultiSP reconstructs denoised data through modality-specific probabilistic decoders, which have been shown to improve the inference of cell-cell communication events. In contrast, SpatialGlue reconstructs reduced-dimensional features (e.g., from principal-component analysis [PCA] or latent semantic indexing [LSI]), limiting its ability to model raw count data and remove noise—a particular advantage for downstream analysis. Unlike methods such as PRESENT that rely heavily on the spatial neighborhood structure, MultiSP is capable of flexibly integrating spatial information and feature similarity according to the characteristics of different tissue structures. To validate the framework design, we conducted ablation studies by removing the VAE component from the omics-specific encoder and adjusting the adversarial alignment loss. Evaluations on both spatial RNA-ADT and RNA-ATAC datasets confirmed that these distinguishing components underlie MultiSP's superior performance (Figure S10).

Different from existing cell-cell communication inference methods, MultiSP infers spatially resolved multimodal cell-cell communication by taking advantage of paired spatial multi-omics data, particularly spatial transcriptomics and proteomics. MultiSP infers cell-cell communication by using gene expression of secreted ligands and protein expression of corresponding surface receptors, which makes biological sense. Moreover, spatial proteomics usually capture the expression of cell-surface proteins instead of secreted proteins due to technical issues. Therefore, MultiSP proposes using the complementary information of the gene and protein modalities, which is different from previous spatial multimodal interaction analysis that uses gene and protein expression to respectively infer gene-gene and protein-protein interactions.[29] By comparing conventional approaches that use mRNA expression to approximate the cell-surface protein expression, we showcase the need for both mRNA and ADT expression in detecting reliable cell-cell communication events.

We conducted a scalability analysis using all studied datasets that have commonly used dataset sizes and types. The running time ranges from 10 s to 3 min, and memory usage ranges from 1 to 17 GB, suggesting the method's practical utility and

(E) Comparison of the normalized mRNA and protein levels of receptors between the inferred cell-cell communication using the denoised RNA-only data, the raw RNA+ADT data, and the denoised RNA+ADT data. *p* values are calculated from a two-sided Wilcoxon rank sum test.
(F) Comparison analysis of COMMOT using denoised RNA-only data versus the proposed method using denoised RNA+ADT data. *p* values are calculated from a two-sided Wilcoxon rank sum test.
(G) Comparison analysis of SpatialDM using denoised RNA-only data versus the proposed method using denoised RNA+ADT data. *p* values are calculated from a two-sided Wilcoxon rank sum test.
(H) The comparison of the inferred IL-16-CD4 signaling network using RNA-only data and RNA+ADT data.
(I) Another example of the inferred H2-AB1-CD4 signaling network. The gene expression of *H2-AB1* was used because it was not measured in ADT.

performance in real-world scenarios (Figures S11A and S11B). To further assess the scalability of MultiSP on larger datasets, we applied it to a mouse spleen dataset generated using the CODEX technology,[81] which contains approximately 80,000 cells and 30 proteins. We created test datasets of varying sizes by randomly sampling subsets of cells. Since this dataset contains only one modality (protein), we generated a second modality by randomly shuffling the expression profiles among the cells. When varying the number of cells from 10,000 to 50,000, the running time of MultiSP is generally under 6 min (Figures S11C and S11D). To handle large datasets, a high-performance computer with greater GPU memory would be beneficial, and the incorporation of graph partitioning strategies and mini-batch strategies such as scNiche[20] would be helpful. We anticipate that MultiSP will be widely used in the fast-developing field of spatial multi-omics technologies and offer new insights into spatially multimodal heterogeneity and cellular communication.

Limitations

Although MultiSP exhibits attractive performance, several avenues for further improvement remain. First, its hybrid approach—using transcriptomic ligands and protein-level receptors—is biologically sound but constrained by the limited coverage of surface proteins in current spatial proteomics platforms. When protein measurements are unavailable, receptor expression could be approximated using transcriptomic data or predicted via cross-omics imputation. Second, although reconstructing raw counts offers advantages over existing methods, computational cost will inevitably increase with larger cell numbers and feature sets. Third, for relatively low-dimensional and dense ADT data, MultiSP currently employs an MSE decoder. Replacing this with a probabilistic decoder (e.g., the MixtureNB decoder from totalVI) resulted in reduced performance (Figure S12), most likely because totalVI is optimized for CITE-seq data, whereas ADT distributions may vary across technologies. Developing improved probabilistic models for ADT data remains an open research question. Finally, MultiSP is designed for spatially aligned multi-omics data profiled within the same cells. Although compatible in principle with both *in-situ*-sequencing- and probe-based platforms, it was primarily evaluated on *in situ* sequencing data due to the scarcity of paired probe-based datasets. As spatial multi-omics evolves, MultiSP could be extended by designing tailored decoders for new data types—for instance, likelihood-based decoders for count-based modalities (e.g., methylation) and MSE-based decoders for data without clear distributions (e.g., high-plex imaging). Further development could also enable the integration of unpaired multi-omics data (e.g., from serial sections or different platforms)[82,83] or the alignment of spatial omics with single-cell multi-omics profiles.[29,30]

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Suoqin Jin (sqjin@whu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The original data used in this paper can be accessed through the following links: (1) SPOTS mouse spleen and breast cancer, GEO: GSE198353; (2) Visium CytAssist human lymph node: https://zenodo.org/records/10362607; (3) MISAR-seq mouse brain: https://www.biosino.org/node/project/detail/OEP003285; (4) spatial ATAC-RNA-seq mouse brain: https://brain-spatial-omics.cells.ucsc.edu/; (5) spatial DBiT-seq mouse embryo, GEO: GSE137986; (6) spatial CITE-seq human skin, GEO: GSE213264; (7) 10× Visium human tonsil: https://www.10xgenomics.com/resources/datasets/gene-protein-expression-library-of-human-tonsil-cytassist-ffpe-2-standard; and (8) CODEX mouse spleen: https://data.mendeley.com/datasets/zjnpwh8m5b/1.
- MultiSP is publicly available as a Python package. Source codes and tutorials have been deposited in GitHub (https://github.com/jinworks/MultiSP) and archived in Zenodo (https://doi.org/10.5281/zenodo.17853075). Codes for reproducing the benchmarking results in this paper are available at https://github.com/jinworks/MultiSP_reproducibility.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The work was supported by a Major Research Plan of the National Natural Science Foundation of China (no. 92374108). The numerical calculations in this paper were done on the supercomputing system at the Supercomputing Center of Wuhan University.

AUTHOR CONTRIBUTIONS

S.J. conceived the project. C.M. and S.J. conducted the research. C.M., S.J., and X.Z. wrote and edited the manuscript. S.J. and X.Z. supervised the research. All authors approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Data preprocessing
  - The MultiSP framework
  - Omics-specific encoder
  - Cross-omics alignment and fusion
  - Omics-specific decoder
  - Model training of MultiSP
  - Implementation details of MultiSP
  - Identification of spatial domains
  - Data denoising
  - Inference of spatially multimodal cell-cell communication
  - Transcription factor regulation analysis
  - Survival analysis
  - Comparison of MultiSP against existing methods
  - Evaluation metrics
  - Dataset description
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2026.101141.

Received: September 14, 2025
Revised: November 14, 2025
Accepted: January 7, 2026
Published: February 5, 2026

REFERENCES

1. Vandereyken, K., Sifrim, A., Thienpont, B., and Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. Nat. Rev. Genet. *24*, 494–515. https://doi.org/10.1038/s41576-025-00899-y.
2. Ben-Chetrit, N., Niu, X., Swett, A.D., Sotelo, J., Jiao, M.S., Stewart, C.M., Potenski, C., Mielinis, P., Roelli, P., Stoeckius, M., and Landau, D.A. (2023). Integration of whole transcriptome spatial profiling with protein markers. Nat. Biotechnol. *41*, 788–793. https://doi.org/10.1038/s41587-022-01536-3.
3. Jiang, F., Zhou, X., Qian, Y., Zhu, M., Wang, L., Li, Z., Shen, Q., Wang, M., Qu, F., Cui, G., et al. (2023). Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. Nat. Methods *20*, 1048–1057. https://doi.org/10.1038/s41592-023-01884-1.
4. Zhang, D., Deng, Y., Kukanja, P., Agirre, E., Bartosovic, M., Dong, M., Ma, C., Ma, S., Su, G., Bao, S., et al. (2023). Spatial epigenome–transcriptome co-profiling of mammalian tissues. Nature *616*, 113–122. https://doi.org/10.1038/s41586-023-05795-1.
5. Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020). High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell *183*, 1665–1681.e18. https://doi.org/10.1016/j.cell.2020.10.026.
6. Liu, Y., DiStasio, M., Su, G., Asashima, H., Enninful, A., Qin, X., Deng, Y., Nam, J., Gao, F., Bordignon, P., et al. (2023). High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. Nat. Biotechnol. *41*, 1405–1409. https://doi.org/10.1038/s41587-023-01676-0.
7. Russell, A.J.C., Weir, J.A., Nadaf, N.M., Shabet, M., Kumar, V., Kambhampati, S., Raichur, R., Marrero, G.J., Liu, S., Balderrama, K.S., et al. (2024). Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. Nature *625*, 101–109. https://doi.org/10.1038/s41586-023-06837-4.
8. Hu, Y., Wan, S., Luo, Y., Li, Y., Wu, T., Deng, W., Jiang, C., Jiang, S., Zhang, Y., Liu, N., et al. (2024). Benchmarking algorithms for single-cell multi-omics prediction and integration. Nat. Methods *21*, 2182–2194. https://doi.org/10.1038/s41592-024-02429-w.
9. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
10. Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K.L., Streets, A., and Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. Nat. Methods *18*, 272–282. https://doi.org/10.1038/s41592-020-01050-x.
11. Tal, A., Gabitto, M.I., and Nir, K.R.V.S.G.-A.M.I.Y. (2023). MultiVI: deep generative model for the integration of multimodal data. Nat. Methods *20*, 1222–1231. https://doi.org/10.1038/s41592-023-01909-9.
12. Yang, X., Mann, K.K., Wu, H., and Ding, J. (2024). scCross: a deep generative model for unifying single-cell multi-omics with seamless integration, cross-modal generation, and in silico exploration. Genome Biol. *25*, 198. https://doi.org/10.1186/s13059-024-03338-z.
13. Jin, S., Zhang, L., and Nie, Q. (2020). scAI: an unsupervised approach for the integrative analysis of parallel single-cell transcriptomic and epigenomic profiles. Genome Biol. *21*, 25. https://doi.org/10.1186/s13059-020-1932-8.
14. Yuan, Z., Zhao, F., Lin, S., Zhao, Y., Yao, J., Cui, Y., Zhang, X.-Y., and Zhao, Y. (2024). Benchmarking spatial clustering methods with spatially resolved transcriptomics data. Nat. Methods *21*, 712–722. https://doi.org/10.1038/s41592-024-02215-8.
15. Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D.J., Lee, E.B., Shinohara, R.T., and Li, M. (2021). SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat. Methods *18*, 1342–1351. https://doi.org/10.1038/s41592-021-01255-8.
16. Dong, K., and Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nat. Commun. *13*, 1739. https://doi.org/10.1038/s41467-022-29439-6.
17. Long, Y., Ang, K.S., Li, M., Chong, K.L.K., Sethi, R., Zhong, C., Xu, H., Ong, Z., Sachaphibulkij, K., Chen, A., et al. (2023). Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. Nat. Commun. *14*, 1155. https://doi.org/10.1038/s41467-023-36796-3.
18. Varrone, M., Tavernari, D., Santamaria-Martínez, A., Walsh, L.A., and Ciriello, G. (2024). CellCharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. Nat. Genet. *56*, 74–84. https://doi.org/10.1038/s41588-023-01588-4.
19. Pham, D., Tan, X., Balderson, B., Xu, J., Grice, L.F., Yoon, S., Willis, E.F., Tran, M., Lam, P.Y., Raghubar, A., et al. (2023). Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. Nat. Commun. *14*, 7739. https://doi.org/10.1038/s41467-023-43120-6.
20. Qian, J., Shao, X., Bao, H., Fang, Y., Guo, W., Li, C., Li, A., Hua, H., and Fan, X. (2025). Identification and characterization of cell niches in tissue from spatial omics data at single-cell resolution. Nat. Commun. *16*, 1693. https://doi.org/10.1038/s41467-025-57029-9.
21. Liu, Y., Li, Z., Chen, X., Cui, X., Gao, Z., and Jiang, R. (2025). INSTINCT: Multi-sample integration of spatial chromatin accessibility sequencing data via stochastic domain translation. Nat. Commun. *16*, 1247. https://doi.org/10.1038/s41467-025-56535-0.
22. Long, Y., Ang, K.S., Sethi, R., Liao, S., Heng, Y., van Olst, L., Ye, S., Zhong, C., Xu, H., Zhang, D., et al. (2024). Deciphering spatial domains from spatial multi-omics with SpatialGlue. Nat. Methods *21*, 1658–1667. https://doi.org/10.1038/s41592-024-02316-4.
23. Tian, T., Zhang, J., Lin, X., Wei, Z., and Hakonarson, H. (2024). Dependency-aware deep generative models for multitasking analysis of spatial omics data. Nat. Methods *21*, 1501–1513. https://doi.org/10.1038/s41592-024-02257-y.
24. Li, Z., Cui, X., Chen, X., Gao, Z., Liu, Y., Pan, Y., Chen, S., Lv, H., Zhai, L., and Jiang, R. (2024). Cross-modality representation and multi-sample integration of spatially resolved omics data. Preprint at bioRxiv. https://doi.org/10.1101/2024.06.10.598155.
25. Zhou, Y., Xiao, X., Dong, L., Tang, C., Xiao, G., and Xu, L. (2025). Cooperative integration of spatially resolved multi-omics data with COSMOS. Nat. Commun. *16*, 27. https://doi.org/10.1038/s41467-024-55204-y.
26. Coleman, K., Schroeder, A., Loth, M., Zhang, D., Park, J.H., Sung, J.-Y., Blank, N., Cowan, A.J., Qian, X., Chen, J., et al. (2025). Resolving tissue complexity by multimodal spatial omics modeling with MISO. Nat. Methods *22*, 530–538. https://doi.org/10.1038/s41592-024-02574-2.
27. Zuo, C., Xia, J., Xu, Y., Xu, Y., Gao, P., Zhang, J., Wang, Y., and Chen, L. (2025). stClinic dissects clinically relevant niches by integrating spatial multi-slice multi-omics data in dynamic graphs. Nat. Commun. *16*, 5317. https://doi.org/10.1038/s41467-025-60575-x.
28. Zhang, T., Zhang, H., Zhao, Z., Shao, S., Jiang, Y., Zhang, X., and Wang, G. (2025). spaMGCN: a graph convolutional network with autoencoder for spatial domain identification using multi-scale adaptation. Genome Biol. *26*, 159. https://doi.org/10.1186/s13059-025-03637-z.
29. Yang, P., Jin, L., Liao, J., Jin, K., Shao, X., Li, C., Qian, J., Cheng, J., Yu, D., Guo, R., et al. (2023). Revealing spatial multimodal heterogeneity in tissues with SpaTrio. Cell Genom. *3*, 100446. https://doi.org/10.1016/j.xgen.2023.100446.
30. Yang, P., Jin, K., Yao, Y., Jin, L., Shao, X., Li, C., Lu, X., and Fan, X. (2025). Spatial integration of multi-omics single-cell data with SIMO. Nat. Commun. *16*, 1265. https://doi.org/10.1038/s41467-025-56523-4.

31. He, S., Bieniosek, M., Song, D., Zhou, J., Chidester, B., Wu, Z., Boen, J., Sharma, P., Trevino, A.E., and Zou, J. (2025). Learning single-cell spatial context through integrated spatial multiomics with CORAL. Preprint at bioRxiv. https://doi.org/10.1101/2025.02.01.636038.

32. Liu, Y., Zou, W., Li, Y., Wang, J., Cai, M., and Cai, H. (2025). Cross-modal Denoising and Integration of Spatial Multi-omics data with CANDIES. Preprint at bioRxiv. https://doi.org/10.1101/2025.04.17.649333.

33. Huang, X., Ma, Z., Meng, D., Liu, Y., Ruan, S., Sun, Q., Zheng, X., and Qiao, Z. (2025). PRAGA: prototype-aware graph adaptive aggregation for spatial multi-modal omics analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, 39(1), 326-333. 39. pp. 326-333. https://doi.org/10.1609/aaai.v39i1.32010.

34. Armingol, E., Baghdassarian, H.M., and Lewis, N.E. (2024). The diversification of methods for studying cell–cell interactions and communication. Nat. Rev. Genet. *25*, 381–400. https://doi.org/10.1038/s41576-023-00685-8.

35. Cheng, X., Huang, H., Su, Y., Nie, Q., Zou, X., and Jin, S. (2025). Cell-cell communication inference and analysis: biological mechanisms, computational approaches, and future opportunities. Preprint at arXiv. https://doi.org/10.48550/arXiv.2512.03497.

36. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088. https://doi.org/10.1038/s41467-021-21246-9.

37. Jin, S., Plikus, M.V., and Nie, Q. (2025). CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. Nat. Protoc. *20*, 180–219. https://doi.org/10.1038/s41596-024-01045-4.

38. Fan, R. (2024). Integrative spatial protein profiling with multi-omics. Nat. Methods *21*, 2223–2225. https://doi.org/10.1038/s41592-024-02533-x.

39. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. Commun. ACM *63*, 139–144. https://doi.org/10.1145/3422622.

40. Kingma, D.P. (2013). Auto-encoding variational bayes. Preprint at arXiv. https://doi.org/10.48550/arXiv.1312.6114.

41. Brombacher, E., Hackenberg, M., Kreutz, C., Binder, H., and Treppner, M. (2022). The performance of deep generative models for learning joint embeddings of single-cell multi-omics data. Front. Mol. Biosci. *9*, 962644. https://doi.org/10.3389/fmolb.2022.962644.

42. Lacan, A., Sebag, M., and Hanczar, B. (2023). GAN-based data augmentation for transcriptomics: survey and comparative assessment. Bioinformatics *39*, i111–i120. https://doi.org/10.1093/bioinformatics/btad239.

43. Mienaltowski, M.J., and Birk, D.E. (2014). Structure, physiology, and biochemistry of collagens. Adv. Exp. Med. Biol. *802*, 5–29. https://doi.org/10.1007/978-94-007-7893-1_2.

44. Betancourt, J., Katzman, S., and Chen, B. (2014). Nuclear factor one B regulates neural stem cell differentiation and axonal projection of corticofugal neurons. J. Comp. Neurol. *522*, 6–35. https://doi.org/10.1002/cne.23373.

45. Lavie, D., Ben-Shmuel, A., Erez, N., and Scherz-Shouval, R. (2022). Cancer-associated fibroblasts in the single-cell era. Nat. Cancer *3*, 793–807. https://doi.org/10.1038/s43018-022-00411-z.

46. Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978. https://doi.org/10.1038/nmeth.4401.

47. Fernandes, C.E.M., Serafim, K.R., Gianlorenço, A.C.L., and Mattioli, R. (2019). Systemic injection of an H4 receptor agonist induces a decrease in CREB and pCREB levels in the cerebellar vermis and prefrontal cortex in mice. Braz. J. Med. Biol. Res. *52*, e8334. https://doi.org/10.1590/1414-431X20198334.

48. Sellers, K., Zyka, V., Lumsden, A.G., and Delogu, A. (2014). Transcriptional control of GABAergic neuronal subtype identity in the thalamus. Neural Dev. *9*, 12–14. https://doi.org/10.1186/1749-8104-9-14.

49. Wang, Y., Dye, C.A., Sohal, V., Long, J.E., Estrada, R.C., Roztocil, T., Lufkin, T., Deisseroth, K., Baraban, S.C., and Rubenstein, J.L.R. (2010). Dlx5 and Dlx6 regulate the development of parvalbumin-expressing cortical interneurons. J. Neurosci. *30*, 5334–5345. https://doi.org/10.1523/JNEUROSCI.5963-09.2010.

50. Sunkin, S.M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T.L., Thompson, C.L., Hawrylycz, M., and Dang, C. (2013). Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Res. *41*, D996–D1008. https://doi.org/10.1093/nar/gks1042.

51. Stanley, G., Gokce, O., Malenka, R.C., Südhof, T.C., and Quake, S.R. (2020). Continuous and discrete neuron types of the adult murine striatum. Neuron *105*, 688–699.e8. https://doi.org/10.1016/j.neuron.2019.11.004.

52. Phillips, R.A., Oh, S., Bach, S.V., Du, Y., Miller, R.A., Kleinman, J.E., Hyde, T.M., Hicks, S.C., Page, S.C., and Martinowich, K. (2025). Transcriptomic characterization of human lateral septum neurons reveals conserved and divergent marker genes across species. iScience *28*, 111820. https://doi.org/10.1016/j.isci.2025.111820.

53. Vincze, A., Mázló, M., Seress, L., Komoly, S., and Ábrahám, H. (2008). A correlative light and electron microscopic study of postnatal myelination in the murine corpus callosum. Int. J. Dev. Neurosci. *26*, 575–584. https://doi.org/10.1016/j.ijdevneu.2008.05.003.

54. Cheung, M., Abu-Elmagd, M., Clevers, H., and Scotting, P.J. (2000). Roles of Sox4 in central nervous system development. Brain Res. Mol. Brain Res. *79*, 180–191. https://doi.org/10.1016/S0169-328X(00)00109-1.

55. Bonate, R., Kurek, G., Hrabak, M., Patterson, S., Padovan-Neto, F., West, A.R., and Steiner, H. (2022). Phosphodiesterase 10A (PDE10A): Regulator of dopamine agonist-induced gene expression in the striatum. Cells *11*, 2214. https://doi.org/10.3390/cells11142214.

56. Tafoya, L.C.R., Mameli, M., Miyashita, T., Guzowski, J.F., Valenzuela, C.F., and Wilson, M.C. (2006). Expression and function of SNAP-25 as a universal SNARE component in GABAergic neurons. J. Neurosci. *26*, 7826–7838. https://doi.org/10.1523/JNEUROSCI.1866-06.2006.

57. Hou, W., Ji, Z., Ji, H., and Hicks, S.C. (2020). A systematic evaluation of single-cell RNA-sequencing imputation methods. Genome Biol. *21*, 218. https://doi.org/10.1186/s13059-020-02132-x.

58. Wang, Y., Song, B., Wang, S., Chen, M., Xie, Y., Xiao, G., Wang, L., and Wang, T. (2022). Sprod for de-noising spatially resolved transcriptomics data based on position and image information. Nat. Methods *19*, 950–958. https://doi.org/10.1038/s41592-022-01560-w.

59. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490. https://doi.org/10.1038/nature14590.

60. Mulvaney, J., and Dabdoub, A. (2012). Atoh1, an essential transcription factor in neurogenesis and intestinal and inner ear development: function, regulation, and context dependency. J. Assoc. Res. Otolaryngol. *13*, 281–293. https://doi.org/10.1007/s10162-012-0317-4.

61. Runge, K., Mathieu, R., Bugeon, S., Lafi, S., Beurrier, C., Sahu, S., Schaller, F., Loubat, A., Herault, L., Gaillard, S., et al. (2021). Disruption of NEUROD2 causes a neurodevelopmental syndrome with autistic features via cell-autonomous defects in forebrain glutamatergic neurons. Mol. Psychiatr. *26*, 6125–6148. https://doi.org/10.1038/s41380-021-01179-x.

62. Zhang, Z., Zhou, J., Tan, P., Pang, Y., Rivkin, A.C., Kirchgessner, M.A., Williams, E., Lee, C.-T., Liu, H., Franklin, A.D., et al. (2021). Epigenomic diversity of cortical projection neurons in the mouse brain. Nature *598*, 167–173. https://doi.org/10.1038/s41586-021-03223-w.

63. Lee, S.-I. (2001). Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I. J. Geogr. Syst. *3*, 369–385. https://doi.org/10.1007/s101090100064.

64. Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R., Lu, Z., Sanger, A., Tokcan, N., Vanderburg, C.R., Segerstolpe, Å., Zhang, M., et al. (2021). Deep learning and alignment of spatially resolved single-cell

transcriptomes with Tangram. Nat. Methods *18*, 1352–1362. https://doi.org/10.1038/s41592-021-01264-7.

65. Krishnan, H., Rayes, J., Miyashita, T., Ishii, G., Retzbach, E.P., Sheehan, S.A., Takemoto, A., Chang, Y.W., Yoneda, K., Asai, J., et al. (2018). Podoplanin: an emerging cancer biomarker and therapeutic target. Cancer Sci. *109*, 1292–1299. https://doi.org/10.1111/cas.13580.
66. Zou, A.E., Kongthong, S., Mueller, A.A., and Brenner, M.B. (2025). Fibroblasts in immune responses, inflammatory diseases and therapeutic implications. Nat. Rev. Rheumatol. *21*, 336–354. https://doi.org/10.1038/s41584-025-01259-0.
67. Martinon, F., Chen, X., Lee, A.-H., and Glimcher, L.H. (2010). Toll-like receptor activation of XBP1 regulates innate immune responses in macrophages. Nat. Immunol. *11*, 411–418. https://doi.org/10.1038/ni.1857.
68. Chen, S., Saeed, A.F.U.H., Liu, Q., Jiang, Q., Xu, H., Xiao, G.G., Rao, L., and Duo, Y. (2023). Macrophages in immunoregulation and therapeutics. Signal Transduct. Targeted Ther. *8*, 207. https://doi.org/10.1038/s41392-023-01452-1.
69. Mantovani, A., Allavena, P., Marchesi, F., and Garlanda, C. (2022). Macrophages as tools and targets in cancer therapy. Nat. Rev. Drug Discov. *21*, 799–820. https://doi.org/10.1038/s41573-022-00520-5.
70. Xie, Z., Zheng, G., Niu, L., Du, K., Li, R., Dan, H., Duan, L., Wu, H., Ren, G., Dou, X., et al. (2025). SPP1+ macrophages in colorectal cancer: Markers of malignancy and promising therapeutic targets. Genes Dis. *12*, 101340. https://doi.org/10.1016/j.gendis.2024.101340.
71. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678. https://doi.org/10.1038/s41587-020-0546-8.
72. Latifkar, A., Wang, F., Mullmann, J.J., Panizza, E., Fernandez, I.R., Ling, L., Miller, A.D., Fischbach, C., Weiss, R.S., Lin, H., et al. (2022). IGF2BP2 promotes cancer progression by degrading the RNA transcript encoding a v-ATPase subunit. Proc. Natl. Acad. Sci. USA *119*, e2200477119. https://doi.org/10.1073/pnas.2200477119.
73. Du, R., Zhang, X., Lu, X., Ma, X., Guo, X., Shi, C., Ren, X., Ma, X., He, Y., Gao, Y., and Liu, Y. (2023). PDPN positive CAFs contribute to HER2 positive breast cancer resistance to trastuzumab by inhibiting antibody-dependent NK cell-mediated cytotoxicity. Drug Resist. Updat. *68*, 100947. https://doi.org/10.1016/j.drup.2023.100947.
74. Tao, J., Shen, L., Zhuang, M., Zhai, C., Zeng, H., Mao, Y., and Liu, X. (2024). Suppression of AGRN enhances CD8+ T cell recruitment and inhibits breast cancer progression. FASEB J. *38*, e23582. https://doi.org/10.1096/fj.202302288R.
75. Seo, E.S., Park, S., Cho, E.Y., Lee, J.E., Jung, H.H., Hyeon, J., An, S., Kim, S.W., Shin, J., Ahn, J.S., et al. (2025). Spatial and genomic profiling of residual breast cancer after neoadjuvant chemotherapy unveil divergent fates for each breast cancer subtype. Cell Rep. Med. *6*, 102164. https://doi.org/10.1016/j.xcrm.2025.102164.
76. 10x Genomics (2023). https://www.10xgenomics.com/resources/datasets/gene-protein-expression-library-of-human-tonsilcytassist-ffpe-2-standard.
77. Lewis, S.M., Williams, A., and Eisenbarth, S.C. (2019). Structure and function of the immune system in the spleen. Sci. Immunol. *4*, eaau6085. https://doi.org/10.1126/sciimmunol.aau6085.
78. Cang, Z., Zhao, Y., Almet, A.A., Stabell, A., Ramos, R., Plikus, M.V., Atwood, S.X., and Nie, Q. (2023). Screening cell–cell communication in spatial transcriptomics via collective optimal transport. Nat. Methods *20*, 218–228. https://doi.org/10.1038/s41592-022-01728-4.
79. Li, Z., Wang, T., Liu, P., and Huang, Y. (2023). SpatialDM for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns. Nat. Commun. *14*, 3995. https://doi.org/10.1038/s41467-023-39608-w.
80. Cruikshank, W., and Little, F. (2008). Interleukin-16: the ins and outs of regulating T-cell activation. Crit. Rev. Immunol. *28*, 467–483. https://doi.org/10.1615/critrevimmunol.v28.i6.10.
81. Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S., and Nolan, G.P. (2018). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. Cell *174*, 968–981.e15. https://doi.org/10.1016/j.cell.2018.07.010.
82. Yan, X., Ang, K.S., van Olst, L., Edwards, A., Watson, T., Zheng, R., Fan, R., Li, M., Gate, D., and Chen, J. (2024). Mosaic integration of spatial multi-omics with SpaMosaic. Preprint at bioRxiv. https://doi.org/10.1101/2024.10.02.616189.
83. Liu, Y., Wang, C., Wang, Z., Chen, L., Li, Z., Song, J., Zou, Q., Gao, R., Qian, B.Z., Feng, X., et al. (2025). High-parameter spatial multi-omics through histology-anchored integration. Nat. Methods. https://doi.org/10.1038/s41592-025-02926-6.
84. Chen, X., Chen, S., Song, S., Gao, Z., Hou, L., Zhang, X., Lv, H., and Jiang, R. (2022). Cell type annotation of single-cell chromatin accessibility data via supervised Bayesian embedding. Nat. Mach. Intell. *4*, 116–126. https://doi.org/10.1038/s42256-021-00432-w.
85. Mulè, M.P., Martins, A.J., and Tsang, J.S. (2022). Normalizing and denoising protein expression data from droplet-based single cell profiling. Nat. Commun. *13*, 2099. https://doi.org/10.1038/s41467-022-29356-8.
86. Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. *19*, 224-12. https://doi.org/10.1186/s13059-018-1603-1.
87. Cao, Z.-J., Wei, L., Lu, S., Yang, D.-C., and Gao, G. (2020). Searching large-scale scRNA-seq databases via unbiased cell embedding with Cell BLAST. Nat. Commun. *11*, 3458. https://doi.org/10.1038/s41467-020-17281-7.
88. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868. https://doi.org/10.1038/nmeth.4380.
89. Qiu, P. (2020). Embracing the dropouts in single-cell RNA-seq analysis. Nat. Commun. *11*, 1169. https://doi.org/10.1038/s41467-020-14976-9.
90. Eraslan, G., Simon, L.M., Mircea, M., Mueller, N.S., and Theis, F.J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. Nat. Commun. *10*, 390. https://doi.org/10.1038/s41467-018-07931-2.
91. Llorens-Bobadilla, E., Zamboni, M., Marklund, M., Bhalla, N., Chen, X., Hartman, J., Frisén, J., and Ståhl, P.L. (2023). Solid-phase capture and profiling of open chromatin by spatial ATAC. Nat. Biotechnol. *41*, 1085–1088. https://doi.org/10.1038/s41587-022-01603-9.
92. Fraley, C., Raftery, A.E., Murphy, T.B., and Scrucca, L. (2012). mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. R J. *8*, 289–317. https://doi.org/10.32614/RJ-2016-021.
93. Kim, H.J., Lin, Y., Geddes, T.A., Yang, J.Y.H., and Yang, P. (2020). CiteFuse enables multi-modal analysis of CITE-seq data. Bioinformatics *36*, 4137–4143. https://doi.org/10.1093/bioinformatics/btaa282.
94. Anthony, M., Oriol, F., Arenillas, D.J., Chih-Yu, C., Grégoire, D., Jessica, L., Wenqiang, S., Casper, S., Ge, T., and Rebecca, W.H. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44*, D110–D115. https://doi.org/10.1093/nar/gkv1176.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Visium CytAssist human lymph node data	Long et al.[22]	https://zenodo.org/records/10362607
SPOTS mouse spleen and breast cancer data	Ben-Chetrit et al.[2]	GSE198353
MISAR-seq mouse brain data	Jiang et al.[3]	https://www.biosino.org/node/project/detail/OEP003285
Spatial ATAC–RNA-seq mouse brain and E13 mouse embryo data	Zhang et al.[4]	https://brain-spatial-omics.cells.ucsc.edu
Spatial CITE-seq human skin data	Liu et al.[6]	GSE213264
Spatial DBiT-seq mouse embryo data	Liu et al.[5]	GSE137986
10× Visium human tonsil gene and protein expression data	10× Genomics	https://www.10xgenomics.com/resources/datasets/gene-protein-expression-library-of-human-tonsil-cytassist-ffpe-2-standard
CODEX mouse spleen data	Goltsev et al.[81]	https://data.mendeley.com/datasets/zjnpwh8m5b/1
Software and algorithms		
R (v4.4.1)	R Core Team	https://www.r-project.org/
Python (v3.11.9)	Python Software Foundation	https://www.python.org/
Pytorch (v 2.4.0)	PyTorch Foundation	https://pytorch.org/
Seurat (v4.3.0)	Hao et al.[9]	https://satijalab.org/seurat/
TotalVI (v1.0.2)	Gayoso et al.[10]	https://github.com/YosefLab/scvi-tools
MultiVI (v1.0.2)	Tal et al.[11]	https://github.com/YosefLab/scvi-tools
stLearn	Pham et al.[19]	https://github.com/BiomedicalMachineLearning/stLearn
STAGATE	Dong et al.[16]	https://github.com/QIFEIDKN/STAGATE
GraphST	Long et al.[17]	https://github.com/JinmiaoChenLab/GraphST
INSTINCT (v1.0.0)	Liu et al.[21]	https://github.com/yyLIU12138/INSTINCT
SpatialGlue	Long et al.[22]	https://github.com/JinmiaoChenLab/SpatialGlue/
PRESENT (v1.0.0)	Li et al.[24]	https://github.com/lizhen18THU/PRESENT
MISO	Coleman et al.[26]	https://github.com/kpcoleman/miso
Sprod	Wang et al.[58]	https://github.com/yunguan-wang/SPROD
spaVAE	Tian et al.[23]	https://github.com/ttgump/spaVAE
CellChat	Jin et al.[36]	https://github.com/jinworks/CellChat
MultiSP	This study	https://github.com/jinworks/MultiSP

METHOD DETAILS

Data preprocessing

For different types of omics data, we apply the corresponding widely used preprocessing methods. For gene expression data (RNA), we first filter out genes expressed in less than 10 cells, then perform library size normalization and log transformation, followed by selecting the top 3,000 highly variable genes. For chromatin accessibility data (ATAC), we filter out peaks accessible in less than 3% cells, followed by frequency-inverse document frequency (TF-IDF) normalization.[84] For protein abundance data (ADT), we apply centered log ratio (CLR) transformation.[85,86]

The MultiSP framework

MultiSP designs a specific variational autoencoder (VAE) with neighborhood structure awareness for each omics, generating low-dimensional embeddings that integrate spatial information for each omics data. These embeddings are then fused across modalities to obtain a unified low-dimensional representation. Beyond spatial location data, MultiSP accepts pairwise combinations or simultaneous input of RNA, ADT, and ATAC modalities. The spatial location data is represented by the spatial coordinate matrix M. The raw

count matrices of RNA, ADT, and ATAC are denoted as X_r, X_p, and X_a respectively, with the preprocessed data matrices represented as $\tilde{X}_r \in R^{n\times d_r}$, $\tilde{X}_p \in R^{n\times d_p}$, and $\tilde{X}_a \in R^{n\times d_a}$ where n denotes the number of cells, d_r represents the number of highly variable genes, d_p represents the number of proteins, and d_a represents the number of filtered peaks. Next, we perform PCA dimensionality reduction on the RNA data $\tilde{X}_r$ to obtainX_{rf}, while LSI is applied for dimensionality reduction of the ATAC data, yielding X_{af}. For the ADT data, let $X_{pf} = \tilde{X}_p$.

We transform M into a spatial proximity graph $G_s(V,E)$, with A_s as the adjacency matrix of the graph, where$a_s(i,j) = 1$ if and only if the Euclidean distance between spots i and j is less than a specific neighbor number k, and 0 otherwise. For RNA and ADT combination data, we choose $k = 3$; for RNA and ATAC combination data, we choose $k = 6$.

Subsequently, we use the KNN algorithm to construct the corresponding feature similarity graphs $G_r(V,E)$, $G_p(V,E)$, and$G_a(V,E)$ based on X_{rf}, X_{pf}, and X_{af}. For each cell, we select k nearest neighbors, where k is set to 20 by default. The corresponding adjacency matrices are denoted as A_r, A_p, and A_a, with a value of 1 indicating that the two cells are neighbors and 0 otherwise. To balance the contributions of the spatial proximity graph and the feature similarity graph, we compute the combined adjacency matrices as follows:

$$A_{r\_combined} = (1 - \lambda)A_s + \lambda A_r \quad \text{(Equation 1)}$$

$$A_{a\_combined} = (1 - \lambda)A_s + \lambda A_a \quad \text{(Equation 2)}$$

$$A_{p\_combined} = (1 - \lambda)A_s + \lambda A_p \quad \text{(Equation 3)}$$

We set the value of λ based on the spatial domain distribution patterns of different tissue types. For tissues exhibiting strong spatial continuity, such as brain tissue, we set $\lambda = 0$ to exclude the feature similarity graph. For other tissue types, we use a default λ value of 0.2 to balance the contributions from spatial and feature-based information. We conducted a sensitivity analysis on the value of k in the spatial KNN graph and the weight of the feature-based KNN graph during graph fusion. The results demonstrate that MultiSP remains robust across a range of these parameter values, confirming the stability and generalizability of the framework (Figure S13).

The fused graphs and reduced-dimension omics feature matrices constitute the input for the core architecture of MultiSP, which is composed of three main modules: the Omics-specific encoder, the Cross-omics alignment and fusion, and the Omics-specific decoder. The details of each module are described below.

Omics-specific encoder

To obtain a neighborhood-aware representation for each cell, we first designed modality-specific Graph Convolutional Network (GCN) encoders for each omics type. The advantage of GCN lies in its ability to capture complex patterns within the graph structure. Specifically, for each omics modality, the preprocessed, dimension-reduced data serve as input, and we derive modality-specific representations based on fused graphs as follows:

$$H_r = \sigma(\tilde{A}_{r\_combined}X_{rf}W_r + b_r) \quad \text{(Equation 4)}$$

$$H_p = \sigma(\tilde{A}_{p\_combined}X_{pf}W_p + b_p) \quad \text{(Equation 5)}$$

$$H_a = \sigma(\tilde{A}_{a\_combined}X_{af}W_a + b_a) \quad \text{(Equation 6)}$$

Here, $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ denotes the normalized adjacency matrix specific to the graph, where D represents the corresponding degree matrix. σ represents the activate function. W and b denote the trainable weight matrices and bias vectors, respectively. The subscripts r, p, a correspond to the RNA, ADT, and ATAC modalities, respectively. This notation applies similarly to the subsequent equations.

Finally, for each modality, we utilize variational autoencoders to encode and obtain the mean H_r^{mean}, $H_p^{mean}$and H_a^{mean}, and the log variance H_r^{var}, H_p^{var} and H_a^{var}. We then perform reparameterization sampling to obtain the latent representations:

$$Z_r = H_r^{mean} + e^{\frac{H_r^{var}}{2}} \times \epsilon \sim \mathcal{N}(0, I) \quad \text{(Equation 7)}$$

$$Z_p = H_p^{mean} + e^{\frac{H_p^{var}}{2}} \times \epsilon \sim \mathcal{N}(0, I) \quad \text{(Equation 8)}$$

$$Z_a = H_a^{mean} + e^{\frac{H_a^{var}}{2}} \times \epsilon \sim \mathcal{N}(0, I) \quad \text{(Equation 9)}$$

Cross-omics alignment and fusion

To integrate low-dimensional representations from different omics that are derived using omics-specific encoders and reside in distinct manifold spaces, we employ generative adversarial networks (GANs) for adversarial alignment. To facilitate subsequent data fusion, we aim to align these representations into a unified manifold space.

We introduce an omics discriminator, which takes as input the latent representations from different omics and assigns labels 0, 1, 2, …, K-1, where K is the total number of omics. The output of the discriminator is a K-dimensional softmax vector, representing the probability distribution over the omics categories. The discriminator is trained to minimize the cross-entropy loss for K-class classification, defined as:

$$Loss_D = -\sum_{i=1}^{n}\sum_{k=1}^{K} y_{i,k}\log\left(\tilde{y}_{i,k}\right) \quad \text{(Equation 10)}$$

where $y_{i,k}$ is the true label for sample i in class k. If sample i belongs to class k, then $y_{i,k} = 1$; otherwise, $y_{i,k} = 0$. $\tilde{y}_{i,k}$ represents the k-th component of the softmax output from the discriminator for sample i. The omics-specific encoders are trained with an opposing objective to fool the discriminator, thus aligning the latent representations from different omics. The corresponding alignment loss function is:

$$Loss_{align} = -Loss_D \quad \text{(Equation 11)}$$

This form of adversarial alignment corresponds to minimizing a generalized Jensen–Shannon divergence among the cell embedding distributions across different omics layers.[87] However, without additional constraints, a purely adversarial objective tends to enforce identical embeddings across modalities for each cell, which may lead to the loss of modality-specific information. To mitigate this, we adopt a weighted adversarial alignment strategy by introducing cell-specific weights in the adversarial loss. This allows the model to retain important modality-specific features:

$$Loss_D = -\sum_{i=1}^{n} w_i \sum_{k=1}^{K} y_{i,k}\log\left(\tilde{y}_{i,k}\right) \quad \text{(Equation 12)}$$

wherew_i are computed based on the cosine similarity between the embeddings of different modalities for each cell, followed by a softmax normalization:

$$w_i = \frac{e^{\sum_{k_i \neq k_j}\cos\left(z_{i,k_i}, z_{i,k_j}\right)}}{\sum_{i=1}^{n} e^{\sum_{k_i \neq k_j}\cos\left(z_{i,k_i}, z_{i,k_j}\right)}} \quad \text{(Equation 13)}$$

This adjustment down-weights the adversarial loss for cells exhibiting greater modality divergence, thereby preventing the potential over-integration and preserving meaningful modality-specific information.

Once the representations are aligned, we fuse them using MLP to obtain a unified latent representation:

$$Z = W_{fus}\left(CONCAT\left(Z_r, Z_p, Z_a\right)\right) + b_{fus} \quad \text{(Equation 14)}$$

where *CONCAT* denotes the concatenation operation, W_{fus} and b_{fus} are trainable weight matrices and bias vectors, respectively.

Omics-specific decoder

There are significant differences in the dimensionality and sparsity of omics data. For instance, protein data (ADT) typically have only a few dozen dimensions, with low sparsity and dropout rates.[88] In contrast, transcriptomics (RNA) and epigenomics (ATAC) data can have tens of thousands to hundreds of thousands of dimensions, exhibiting higher levels of dropout and extreme sparsity.[89] As a result, we designed specific decoders for each omics type.

For RNA count data X_r, following prior single-cell studies,[90] we modeled the data using a zero-inflated negative binomial (ZINB) distribution:

$$X_r^{i,j} \sim ZINB\left(X_r^{i,j}; \pi_{i,j}, \mu_{i,j}, \theta_{i,j}\right) \quad \text{(Equation 15)}$$

$$ZINB\left(X_r^{i,j}; \pi_{i,j}, \mu_{i,j}, \theta_{i,j}\right) = \pi_{i,j}\delta_0\left(X_r^{i,j}\right) + \left(1 - \pi_{i,j}\right)NB\left(X_r^{i,j}; \mu_{i,j}, \theta_{i,j}\right) \quad \text{(Equation 16)}$$

$$\delta_0\left(X_r^{i,j}\right) = \begin{cases} 1, if\ X_r^{i,j}\ equals\ 0 \\ 0, otherwise \end{cases} \quad \text{(Equation 17)}$$

$$NB\left(X_r^{i,j};\mu_{i,j},\theta_{i,j}\right)=\frac{\Gamma\left(X_r^{i,j}+\theta_{i,j}\right)}{\theta_{i,j}}\left(\frac{\theta_{i,j}}{\theta_{i,j}+\mu_{i,j}}\right)^{\theta_{i,j}}\left(\frac{\mu_{i,j}}{\theta_{i,j}+\mu_{i,j}}\right)^{X_r^{i,j}} \quad \text{(Equation 18)}$$

where i and j denote the i-th cell and the j-th gene, and π, μ, θ represent the zero-inflation rate, mean, and dispersion parameter, respectively. These parameters are learned through MLP:

$$\pi=sigmoid\left(f_{r,\pi}^{MLP}(Z_r)\right),\mu=diag(S_r)\times exp\left(f_{r,\mu}^{MLP}(Z_r)\right) \quad \text{(Equation 19)}$$

$$\theta=softplus\left(f_{r,\theta}^{MLP}(Z_r)\right) \quad \text{(Equation 20)}$$

where *sigmoid*, *exp*, and *softplus* denote activation functions, and S_r is a scaling factor used for library size normalization. The loss function for the ZINB decoder is defined as:

$$L_{ZINB}=\frac{1}{nd_r}\sum_{i=1}^{n}\sum_{j=1}^{d_r}\left(-log\,ZINB\left(X_r^{i,j};\pi_{i,j},\mu_{i,j},\theta_{i,j}\right)+\gamma\pi_{i,j}^2\right) \quad \text{(Equation 21)}$$

The term $\gamma\pi_{i,j}^2$ acts as a ridge prior on the dropout probabilities and accounts for zero inflation. The regularization parameter γ controls the weight of the dropout process. In our model, we set $\gamma = 0.5$ by default.

For ATAC count data X_a, inspired by single-cell research,[91] we modeled the data using a zero-inflated Poisson (ZIP) distribution:

$$X_a^{i,j}\sim ZIP\left(X_a^{i,j};\pi_{i,j},\lambda_{i,j}\right) \quad \text{(Equation 22)}$$

$$ZIP\left(X_a^{i,j};\pi_{i,j},\lambda_{i,j}\right)=\pi_{i,j}\delta_0\left(X_a^{i,j}\right)+\left(1-\pi_{i,j}\right)Possion\left(X_a^{i,j};\lambda_{i,j}\right) \quad \text{(Equation 23)}$$

$$\delta_0\left(X_a^{i,j}\right)=\begin{cases}1, if\ X_a^{i,j}\ equals\ 0\\ 0, otherwise\end{cases} \quad \text{(Equation 24)}$$

$$Possion\left(X_a^{i,j};\lambda_{i,j}\right)=\frac{\lambda_{i,j}^{X_a^{i,j}}e^{-\lambda_{i,j}}}{X_a^{i,j}!} \quad \text{(Equation 25)}$$

where i and j represent the i-th cell and j-th peak, π and λ denote the zero-inflation rate and mean parameter, respectively. These parameters are learned through MLP:

$$\pi=sigmoid\left(f_{a,\pi}^{MLP}(Z_a)\right),\rho=\left(f_{a,\rho}^{MLP}(Z_a)\right) \quad \text{(Equation 26)}$$

$$\lambda=diag(S_a)\times softmax(\rho+r^c) \quad \text{(Equation 27)}$$

where *sigmoid* and *softmax* are activation functions, S_a represents the library size for each cell, and r^c captures peak-specific bias. The loss function for the ZIP decoder is defined as:

$$L_{ZIP}=\frac{1}{nd_a}\sum_{i=1}^{n}\sum_{j=1}^{d_a}\left(-log\,ZIP\left(X_a^{i,j};\pi_{i,j},\lambda_{i,j}\right)+\gamma\pi_{i,j}^2\right) \quad \text{(Equation 28)}$$

Similar to the ZINB decoder, the regularization parameter γ was set to 0.5 by default.

Finally, for ADT data, we reconstruct the normalized feature matrix directly using MLP as follows:

$$\tilde{X}_{pf}=f_p^{MLP}(Z_p) \quad \text{(Equation 29)}$$

and the corresponding mean squared error (MSE) loss function is:

$$L_{MSE}=\frac{1}{nd_p}\sum_{i=1}^{n}\sum_{j=1}^{d_p}\left(X_{pf}^{i,j}-\tilde{X}_{pf}^{i,j}\right)^2 \quad \text{(Equation 30)}$$

Model training of MultiSP

The model training process is divided into two stages. In the first stage, each modality's variational autoencoder (VAE) is pretrained independently, with the corresponding loss functions defined as follows:

$$Loss_r=\lambda_{kl}KL\left(\mathcal{N}\left(H_r^{mean},H_r^{var}\right)\|\mathcal{N}(0,I)\right)+\lambda_{zinb}L_{ZINB} \quad \text{(Equation 31)}$$

$$Loss_a = \lambda_{kl} KL\left(\mathcal{N}\left(H_a^{mean}, H_a^{var}\right) \| \mathcal{N}(0, I)\right) + \lambda_{zip} L_{ZIP} \quad \text{(Equation 32)}$$

$$Loss_p = \lambda_{kl} KL\left(\mathcal{N}\left(H_p^{mean}, H_p^{var}\right) \| \mathcal{N}(0, I)\right) + \lambda_{mse} L_{MSE} \quad \text{(Equation 33)}$$

where KL denotes the Kullback-Leibler divergence loss, and λ_{kl}, λ_{zinb}, λ_{zip} and λ_{mse} are the weighting coefficients for the corresponding losses. The values are set as $\lambda_{kl} = 0.005$ and $\lambda_{zinb} = \lambda_{zip} = \lambda_{mse} = 1$.

The first stage allows each encoder to learn an initial low-dimensional embedding specific to its corresponding modality. Essentially, this pretraining step serves a similar purpose to initializing the generator in a GAN framework and helps mitigate instability before adversarial optimization begins.

In the second stage, an omics discriminator is introduced. During the decoding process, the unified representation Z fusing information from multiple modalities replaces the modality-specific latent representations Z_r, Z_a and Z_p. The discriminator and each modality's VAE are alternately optimized. The discriminator's loss function is denoted as $Loss_D$, and the overall loss function combining all the VAEs is defined as:

$$Loss = \lambda_r Loss_r + \lambda_a Loss_a + \lambda_p Loss_p + \lambda_{align} Loss_{align} \quad \text{(Equation 34)}$$

where λ_r、λ_a、λ_p and λ_{align} are the corresponding loss weights. By default, we set $\lambda_{align} = 0.1$. The sensitivity analysis of this weighting coefficient for the adversarial loss demonstrates that MultiSP remains robust across a range of this parameter (Figure S13). For all the RNA-ADT datasets, the corresponding weights λ_r *and* $\lambda_p$were both set to 1; For the MISAR-seq dataset, the corresponding weights λ_r *and* λ_a were set to 1 and 5; For the Spatial ATAC-RNA-seq dataset, the corresponding weights λ_r *and* $\lambda_a$were set to 1 and 10.

Implementation details of MultiSP

We performed dimensionality reduction on RNA and ATAC data using PCA and LSI, respectively, reducing them to 100 dimensions. The corresponding dimensional parameters for the GCN encoder were set to (100, 128) for RNA and ATAC data, while for ADT data, the GCN encoder dimensions were (d_p, 128). The input and output dimensions of the MLP that encodes the mean and log variance were both set to 128. The dimensional parameters for the ZINB decoder were set to (128, 1024, d_r), for the ZIP decoder to (128, 1024, d_a), and for the ADT decoder to (128, 256, d_p). The variational autoencoder (VAE) was trained using the Adam optimizer. To improve stability, the discriminator was trained using the momentum-free RMSprop optimizer. We conducted 100 epochs during the initial pretraining stage for all datasets. In the second training stage, we performed 50 epochs for RNA–ADT data and 100 epochs for RNA–ATAC data. The learning rate of the Adam and RMSprop optimizers were set to 0.001 and 0.005, respectively. For the P22 mouse brain data, we accelerated the training by utilizing an MLP decoder instead of the ZIP decoder to directly reconstruct the features of LSI dimensionality reduction. For the human tonsil dataset, Hematoxylin and Eosin (H&E)-stained images were treated as an additional modality. Preprocessed image features from MISO[26] were integrated into the MultiSP architecture and reconstructed using a standard MLP decoder. Both the initial pretraining stage and the subsequent training stage were conducted for 300 epochs.

While the low-dimensional embeddings for each cell are modeled using a normal distribution through variational autoencoders (VAEs), MultiSP employs a reparameterization technique during training to enable backpropagation through the stochastic layer. Specifically, embedding vectors are sampled from the learned distributions and used as inputs to the decoder. However, during inference, we use the mean of each cell's latent distribution as its final representation. This approach ensures that, once the model is trained and the random seed is fixed, all downstream analyses are deterministic.

Identification of spatial domains

By default, spatial domains were identified using the mclust[92] clustering algorithm, applied to the joint latent representations generated by MultiSP. For labeled datasets, we set the number of clusters to be the same as the ground truth. For unlabeled datasets, we either set the number of clusters to be the same as the previous studies or determine the number of clusters based on domain-specific differential expression patterns such as in the analysis of mouse spleen and breast cancer data. In addition, we also applied both Louvain and Leiden clustering algorithms within the MultiSP framework and found that MultiSP outperformed competing methods under these settings as well (Figures S14A and S14B). We also incorporated Louvain and Leiden as optional clustering choices in the Python package to enhance flexibility and compatibility with standard practices.

Data denoising

For RNA data, we use the output $\tilde{\mu} = exp\left(f_{r,\mu}^{MLP}(Z_r)\right)$ from Equation 19 as the denoised gene counts, which removes the effect of library size. To obtain more robust estimates, $\tilde{\mu}$ is calculated based on H_r^{mean} (Equation 7), which represents the mean of Z_r. For ATAC data, we use the output $\rho = \left(f_{a,\rho}^{MLP}(Z_a)\right)$ from Equation 26 as the denoised peak counts, which both correct library size effects and peak-specific bias. For ADT data,we directly use $\tilde{X}_{pf}$ from Equation 29 as the denoised, normalized expression values.

Inference of spatially multimodal cell-cell communication

We inferred spatially multimodal cell-cell communication by extending our previous cell-cell communication inference method CellChat and utilizing denoised spatial multi-omics data that simultaneously captures transcriptomics and proteomics. Because

protein data often exhibit technical biases such as background signals arising from ambient or nonspecifically bound antibodies, we set low protein expression values to zero to mitigate these effects. Similar to a previous study,[93] we defined the expression values of proteins less than 0.5 as the low-expressions. To account for the different scales of gene and protein expression, we normalized each feature to a range of 0–1. We observed consistent results when varying the threshold for the low-expression values of proteins (Figure S7B). We first identified pairs of cells that were physically close to one another to have biologically realistic interactions. For the secreting signaling, we used the maximal possible molecular interaction range (by default 250 μm); for the contact-dependent signaling, we required that communicating cells were in direct contact with each other. Secondly, we identified combinations of spatial domains that have enough nearby cell-cell pairs. Thirdly, based on the CellChat's framework, we defined an interaction score by calculating the average gene expression of a ligand in a spatial domain and the average protein expression of a surface receptor in another spatial domain using 10% truncated mean. Finally, the significant interactions were identified using a permutation test by randomly permuting the group labels of cells and the interactions between spatially proximal spatial domains with p-value <0.05 were considered significant. When receptors are not measured at the protein level, users can either use the gene expression as a proxy or predict the corresponding protein expression using cross-omics imputation methods.

Transcription factor regulation analysis

We performed transcription factor motif enrichment analysis and estimated transcription factor activity using the R package 'chromVAR'. The positional weight matrix used in the process was obtained from the JASPAR2016 database.[94] chromVAR calculates the bias corrected deviations in accessibility. For each motif, there is a value for each cell, which measures how different the accessibility for loci with that motif is from the expected accessibility based on the average of all the cells. We computed the average gene expression levels and transcription factor activities for all cells across various spatial domains. These values were normalized to a range of [0, 1], enabling a direct comparison.

Survival analysis

Survival analysis was performed using R package 'survival'. Patients were classified into a high group and a low group using the 'surv_cutpoint' function from R package 'survminer' based on the computed GSVA scores. The survival curve of patients is then calculated using the Kaplan–Meier method in 'survival' package and visualized using the 'ggsurvplot' function from R package 'survminer'. The survival curve difference of patients between different groups is evaluated by two-sided log rank test (Mantel–Cox test) without multiple testing correction.

Comparison of MultiSP against existing methods

We evaluate MultiSP's ability in spatial domain detection against other twelve methods, including three methods designed for integrating single-cell multi-omics data (Seurat v4(WNN), TotalVI and MultiVI), four methods designed for integrating spatial single omics data (STAGATE, GraphST, INSTINCT and stLearn), three methods designed for integrating spatial multi-omics data (SpatialGlue, PRESENT and MISO), as well as two baseline methods that only use one modality (denoted by RNA and Protein (or ATAC) respectively). Of note, TotalVI and MultiVI, implemented by scVI package (version 1.0.2), are used for the integrative analysis of spatial RNA-ADT and RNA-ATAC omics data, respectively. The two baseline methods are obtained by creating a simplified version of MultiSP framework that takes only one modality data as input.

In addition, we compared MultiSP with spatial data denoising methods, including Sprod and spaVAE, and cell-cell communication inference methods, including COMMOT and SpatialDM. For both COMMOT and SpatialDM, we did not normalize each gene feature to a range of 0–1, as such normalization is not required in the corresponding tutorials. In SpatialDM, weight matrix was calculated using 'spatialdm.weight_matrix' with options l = 2.5, cutoff = 0.

We ran these methods with suggested parameter settings by following the corresponding tutorials. Due to the randomness of deep learning-based methods, we perform repeated experiments using different random seeds and conducted statistical validation. Specifically, for human lymph node dataset, we carry out 10 independent runs, each corresponding to a distinct random seed, and reported the variation in the performance metrics using box-plots. To assess statistical significance in performance differences across methods, we conduct Wilcoxon rank-sum test for the results obtained across multiple runs with different seeds. p-values from Wilcoxon rank-sum tests are indicated on the box-plots.

Evaluation metrics

To evaluate the spatial domain identification performance of different methods, we utilized five widely used supervised clustering evaluation metrics, including Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), Entropy, Purity and F1 score and four unsupervised metrics that capture spatial continuity in the previous benchmarking study,[14] including CHAOS, PAS, ASW, and Moran's I. Better clustering performance corresponds to higher values for NMI, ARI, Purity, F1 score and lower values for Entropy. Lower values of CHAOS and PAS and higher values of ASW and Moran's I indicate better performance in capturing spatial continuity. We also calculated both supervised and unsupervised overall scores by combining the results from all corresponding evaluation metrics. Taking the supervised overall score as an example, we first calculated the five scores for each benchmarking method, then sorted the ARI, NMI, Purity and F1 score values of different methods in ascending order to get $rank_{ARI}$, $rank_{NMI}$, $rank_{Purity}$

and $rank_{F1score}$, and sorted the Entropy values in descending order to get $rank_{Entropy}$. Finally, we calculated the average value to obtain the overall score value as follows:

$$Overall\ score(\text{supervised}) = \frac{1}{5}\left(\begin{array}{c} rank_{ARI} + rank_{NMI} + rank_{Entropy} \\ + rank_{Purity} + rank_{F1\ score} \end{array}\right) \quad \text{(Equation 35)}$$

The overall score of unsupervised metrics was similarly calculated as follows:

$$Overall\ score(\text{unsupervised}) = \frac{1}{4}(rank_{CHAOS} + rank_{PAS} + rank_{ASW} + rank_{Moran's\ I}) \quad \text{(Equation 36)}$$

Dataset description

Visium CytAssist human lymph node data

The human lymph node dataset is generated by the Visium CytAssist Spatial Gene and Protein Expression platform (10× Genomics). The dataset contains the spatial coordinates and the expression level of 18,085 genes and 31 proteins across 3,484 cells. The ground truth labels were obtained from the SpatialGlue study, where tissue domains were annotated by clinical experts based on Hematoxylin and Eosin (H&E) stained sections.

SPOTS mouse spleen and breast cancer data

The mouse spleen and breast cancer data are generated by a high-throughput spatial transcriptomics and proteomics co-profiling technology, i.e., Spatial PrOtein and Transcriptome Sequencing (SPOTS). The mouse spleen dataset provides the spatial coordinates and the expression level of 32,285 genes and 21 proteins across 2,568 cells. The mouse breast cancer dataset provides the spatial coordinates and the expression level of 32,286 genes and 32 proteins across 1,978 cells.

MISAR-seq mouse brain data

The MISAR-seq dataset comprises eight spATAC-seq samples and their corresponding spatial transcriptomics (SRT) profiles from E11.0, E13.5, E15.5 and E18.5 developmental stages. The four samples include E11_0-S1(32,285 genes and 242,024 peaks across 1,258 cells), E13_5-S1(32,285 genes and 271,126 peaks across 1,777 cells), E15_5-S1(32,285 genes and 265,014 peaks across 1,949 cells, and E18_5-S1(32,285 genes and 294,734 peaks across 2,129 cells. We used anatomical annotations available in the original publication of the MISAR-seq technology. These annotations were derived from spatial alignment to corresponding histological images and referenced against standard mouse brain atlases.

Spatial ATAC-RNA-seq mouse brain and E13 mouse embryo data

The mouse brain datasets includes three samples of mouse postnatal day 21/22 (P21/22). Microfluidic barcoding was used to capture spatial location and combined with *in situ* Tn5 transposition chemistry to capture chromatin accessibility. We used the Mouse P22 spatial RNA+ATAC dataset which provides the spatial coordinates and the expression level of 22,914 genes and the accessibility of 121,068 peaks across 9,215 cells. In experiments conducted on embryonic day 13 (E13) mouse embryos, with a pixel size of 50 μm, the E13 mouse embryo dataset comprises 2,186 cells, 20,900 genes, and a peak count reaching up to 87,173.

Spatial CITE-seq human skin data

Spatial CITE-seq was employed to map early immune cell activation within a skin biopsy tissue collected from the injection site of a Coronavirus Disease 2019 (COVID-19) mRNA vaccine, co-profiling spatially resolved proteome and transcriptome. Spatial CITE-seq employs a cocktail of approximately 200–300 antibody-derived tags (ADTs) to stain a tissue slide. This dataset comprises 15,486 genes and 283 proteins on 1,691 cells.

Spatial DBiT-seq mouse embryo data

We used the '0713aL' and '0713cL' samples, which are the sequencing data for the mouse embryonic brain region. The dataset contains 1,789 cells with 18,787 genes and 22 proteins.

10× visium human tonsil gene and protein expression data

The dataset contains 4,191 cells with 18,085 genes and 35 proteins. The annotation of germinal center was provided by the authors of MISO.[26]

CODEX mouse spleen data

We used the sample 'BALBc-1' for our scalability analysis. The dataset contains 82,251 cells and 30 proteins.

QUANTIFICATION AND STATISTICAL ANALYSIS

The quantitative and statistical analyses are described in the relevant sections of the method details and in the figure legends. R (version 4.4.1) and Python 3.11.9 were used for all statistical analyses.